BIRCH, STEWART, KOLASCH & BIRCH, LLP

INTELLECTUAL PROPERTY LAW
8110 GATEHOUSE ROAD
SUITE 500 EAST
FALLS CHURCH, VA 22042-1210
USA
(703) 205-8000

FAX (703) 205-8050
(703) 698-8590 (G IV)

e-mail: mailroom@bskb.com
web: http://www.bskb.com

CALIFORNIA OFFICE
COSTA MESA, CALIFORNIA

TERRELL C BIRCH
RAYMOND C STEWART
JOSEPH A KOLASCH
JAMES M SLATTERY
BERNARD L SWEENEY*
MICHAEL K MUTTER
CHARLES GORENSTEIN
GERALD M MURPHY, JR
LEONARD R SVENSSON
TERRY L CLARK
ANDREW D MEIKLE
MARC S WEINER
JOE McKINNEY MUNCY
ROBERT J KENNEY
DONALD J DALEY
JOHN W BAILEY
JOHN A CASTELLANO, III
GARY D YACURA

OF COUNSEL
HERBERT M BIRCH (1905-1996)
ELLIOT A GOLDBERG*
WILLIAM L GATES*
EDWARD H VALANCE
RUPERT J BRADY (RET)*
F PRINCE BUTLER
FRED S WHISENHUNT

*ADMITTED TO A BAR OTHER THAN VA

THOMAS S AUCHTERLONIE
JAMES T ELLER, JR
SCOTT L LOWE
MARK J NUELL, Ph D
D RICHARD ANDERSON
PAUL C LEWIS
MARK W MILSTEAD*
RICHARD J GALLAGHER
JAYNE M SAYDAH*

REG PATENT AGENTS
FREDERICK R HANDREN
MARYANNE ARMSTRONG, Ph D
MAKI HATSUMI
MIKE S RYU
CRAIG A McROBBIE
GARTH M DAHLEN, Ph D
LAURA C LUTZ
ROBERT E GOOZNER, Ph D
HYUNG N SOHN
MATTHEW J LATTIG
ALAN PEDERSEN-GILES
C KEITH MONTGOMERY
TIMOTHY R WYCKOFF
KRISTI L RUPERT, Ph D
LARRY J HUME
ALBERT LEE
HRAYR A SAYADIAN, Ph D

Date: October 13, 2000

Docket No.: 2750-1235P

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

Sir:

As authorized by the inventor(s), transmitted herewith for filing is a patent application applied for on behalf of the inventor(s) according to the provisions of 37 C.F.R. § 1.41(c), which claims priority under 35 U.S.C. § 119(e) of Provisional Application No. 60/159,330 filed on October 14, 1999.

Inventor(s): Nickolai ALEXANDROV, Vyacheslav BROVER

For: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES ENCODED THEREBY

Enclosed are:

☒ A specification consisting of a Description (1078 pages), Table 1 (909 pages), Table 2 (2477 pages), Claims (5 pages), Schematic (1 page), and an Abstract (1 page) totaling four-thousand four-hundred and seventy-one (4471) pages

☐ ( ) sheet(s) of formal drawings

☐ Certified copy of Priority Document(s)

☒ Executed Declaration in accordance with 37 C.F.R. § 1.64 will follow

☒ A statement to establish small entity status under 37 C.F.R. § 1.9 and 37 C.F.R. § 1.27

☐ Preliminary Amendment

☒ Information Sheet

☐ Information Disclosure Statement, PTO-1449 and reference(s)

☐ Amend the specification by inserting before the first line the sentence:

--This application claims priority on provisional Application No. filed on , the entire contents of which are hereby incorporated by reference.--

☒ Other: Power of Attorney regarding Small Entity Statement, ATCC Deposit receipts PTA-595, PTA-1161, PTA-1411, CD containing Specification

The filing fee has been calculated as shown below:

			LARGE ENTITY	**SMALL ENTITY**
	BASIC FEE		**$690.00**	**$345.00**
	NUMBER FILED	NUMBER EXTRA	RATE FEE	RATE FEE
TOTAL CLAIMS	50- 20 =	30	X 18 = $0.00	x 9 = 270
INDEPENDENT CLAIMS	5- 3 =	2	x 78 = $0.00	x 39 = 78
☐ MULTIPLE DEPENDENT CLAIMS PRESENTED			+ $260.00	+ $130.00
		TOTAL	$0.00	$693.00

☒ The application transmitted herewith is filed in accordance with 37 C.F.R. § 1.41(c). The undersigned has been authorized by the inventor(s) to file the present application. The original duly executed declaration together with the surcharge will be forwarded in due course.

☒ A check in the amount of $693.00 to cover the filing fee is enclosed.

Docket No. 2750-1235P

☐ Please charge Deposit Account No. 02-2448 in the amount of $0.00. A triplicate copy of this transmittal form is enclosed.

☒ Please send correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP **or** Customer No. 2292
P.O. Box 747
Falls Church, VA 22040-0747
Telephone: (703) 205-8000

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. §§ 1.16 or 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By Raymond C. Stewart
Raymond C. Stewart, #21,066

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

RCS/CAV
2750-1235P

Attachments

(Rev. 04/19/2000)

STATEMENT CLAIMING SMALL ENTITY STATUS
(37 CFR 1.9(f) & 1.27(c)) - SMALL BUSINESS CONCERN

Docket Number: 2750-1235P

Applicant, Patentee, or Identifier: N. ALEXANDROV et al.
Application or Patent No.: NEW Patent Application
Filed or Issued: October 13, 2000
Title: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES ENCODED THEREBY

I hereby state that I am

☐ the owner of the small business concern identified below:
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN CERES, INC.
ADDRESS OF SMALL BUSINESS CONCERN 3007 Malibu Canyon Road Malibu, CA 90265

I hereby state that the above identified small business concern qualifies as a small business concern as defined in 37 CFR Part 121 for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby state that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention described in:

☒ the specification filed herewith with title as listed above.
☐ the application identified above.
☐ the patent identified above.

If the rights held by the above identified small business concern are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d), or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:

☒ no such person, concern, or organization exists.
☐ each such person, concern, or organization is listed below.

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is not longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING Raymond C. Stewart (Reg. No. 21,066)

TITLE IN ORGANIZATION OF PERSON SIGNING Legal Representative of CERES, INC.

ADDRESS OF PERSON SIGNING Birch, Stewart, Kolasch and Birch, LLP.

P.O. Box 747 Falls Church, VA 22040-0747

SIGNATURE Raymond C. Stewart DATE October 13, 2000

Rev. 10/12/1998

PATENT
2750-1235P

IN THE U.S. PATENT AND TRADEMARK OFFICE

INFORMATION SHEET

Applicant: Nickolai ALEXANDROV and Vyacheslav BROVER

Appl. No.: NEW

Filed: October 13, 2000

For: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES ENCODED THEREBY

Priority Claimed: 2750-0573P 60/159,330 October 14, 1999

Send Correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP or **CUSTOMER NO. 2292**
P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

The above information is submitted to advise the U.S.P.T.O. of all relevant facts in connection with the present application.

A timely executed Declaration in accordance with 37 C.F.R. § 1.64 will follow.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By Raymond C. Stewart
Raymond C. Stewart, #21,066

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

RCS/CAV
2750-1235P

(Rev. 04/19/2000)

SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES ENCODED THEREBY

This application claims priority under 35 USC §119(e), §119(a-d) and §120 of the following applications, the entire contents of which are hereby incorporated by reference:

Country	Filing Date	Attorney No.	Client No.	Application No.
United States	10/14/1999	2750-0573P	80145.001	60/159,330

FIELD OF THE INVENTION

The present invention relates to isolated polynucleotides that represent a complete gene, or a fragment thereof, that is expressed. In addition, the present invention relates to the polypeptide or protein corresponding to the coding sequence of these polynucleotides. The present invention also relates to isolated polynucleotides that represent regulatory regions of genes. The present invention also relates to isolated polynucleotides that represent untranslated regions of genes. The present invention further relates to the use of these isolated polynucleotides and polypeptides and proteins.

DESCRIPTION OF THE RELATED ART

Efforts to map and sequence the genome of a number of organisms are in progress; a few complete genome sequences, for example those of *E. coli* and *Saccharomyces cerevisiae* are known (Blattner et al., *Science* 277:1453 (1997); Goffeau et al., *Science* 274:546 (1996)). The complete genome of a multicellular organism, *C. elegans*, has also been sequenced (See, the *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998)). To date, no complete genome of a plant has been sequenced, nor has a complete cDNA complement of any plant been sequenced.

SUMMARY OF THE INVENTION

The present invention comprises polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, fragments of genes, and/or regulatory elements of genes and/or regions with other functions and/or intergenic regions, hereinafter collectively referred to as Sequence-Determined DNA Fragments (SDFs), from different plant species, particularly corn, wheat, soybean, rice and *Arabidopsis thaliana,* and other plants and or mutants, variants, fragments or fusions of said SDFs and polypeptides or proteins derived therefrom. In some instances, the SDFs span the entirety of a protein-coding segment. In some instances, the entirety of an mRNA is represented. Other objects of the invention that are also represented by SDFs of the invention are control sequences, such as, but

not limited to, promoters. Complements of any sequence of the invention are also considered part of the invention.

Other objects of the invention are polynucleotides comprising exon sequences, polynucleotides comprising intron sequences, polynucleotides comprising introns together with exons, intron/exon junction sequences, 5' untranslated sequences, and 3' untranslated sequences of the SDFs of the present invention. Polynucleotides representing the joinder of any exons described herein, in any arrangement, for example, to produce a sequence encoding any desirable amino acid sequence are within the scope of the invention.

The present invention also resides in probes useful for isolating and identifying nucleic acids that hybridize to an SDF of the invention. The probes can be of any length, but more typically are 12-2000 nucleotides in length; more typically, 15 to 200 nucleotides long; even more typically, 18 to 100 nucleotides long.

Yet another object of the invention is a method of isolating and/or identifying nucleic acids using the following steps:

(a) contacting a probe of the instant invention with a polynucleotide sample under conditions that permit hybridization and formation of a polynucleotide duplex; and

(b) detecting and/or isolating the duplex of step (a).

The conditions for hybridization can be from low to moderate to high stringency conditions. The sample can include a polynucleotide having a sequence unique in a plant genome. Probes and methods of the invention are useful, for example, without limitation, for mapping of genetic traits and/or for positional cloning of a desired fragment of genomic DNA.

Probes and methods of the invention can also be used for detecting alternatively spliced messages within a species. Probes and methods of the invention can further be used to detect or isolate related genes in other plant species using genomic DNA (gDNA) and/or cDNA libraries. In some instances, especially when longer probes and low to moderate stringency hybridization conditions are used; the probe will hybridize to a plurality of cDNA and/or gDNA sequences of a plant. This approach is useful for isolating representatives of gene families which are identifiable by possession of a common functional domain in the gene product or which have common cis-acting regulatory sequences. This approach is also useful for identifying orthologous genes from other organisms.

The present invention also resides in constructs for modulating the expression of the genes comprised of all or a fragment of an SDF. The constructs comprise all or a fragment of the expressed SDF, or of a complementary sequence. Examples of constructs include ribozymes comprising RNA encoded by an SDF or by a sequence complementary thereto,

antisense constructs, constructs comprising coding regions or parts thereof, constructs comprising promoters, introns, untranslated regions, scaffold attachment regions, methylating regions, enhancing or reducing regions, DNA and chromatin conformation modifying sequences, etc. Such constructs can be constructed using viral, plasmid, bacterial artificial chromosomes (BACs), plasmid artificial chromosomes (PACs), autonomous plant plasmids, plant artificial chromosomes or other types of vectors and exist in the plant as autonomous replicating sequences or as DNA integrated into the genome. When inserted into a host cell the construct is, preferably, functionally integrated with, or operatively linked to, a heterologous polynucleotide. For instance, a coding region from an SDF might be operably linked to a promoter that is functional in a plant.

The present invention also resides in host cells, including bacterial or yeast cells or plant cells, and plants that harbor constructs such as described above. Another aspect of the invention relates to methods for modulating expression of specific genes in plants by expression of the coding sequence of the constructs, by regulation of expression of one or more endogenous genes in a plant or by suppression of expression of the polynucleotides of the invention in a plant. Methods of modulation of gene expression include without limitation (1) inserting into a host cell additional copies of a polynucleotide comprising a coding sequence; (2) modulating an endogenous promoter in a host cell; (3) inserting antisense or ribozyme constructs into a host cell and (4) inserting into a host cell a polynucleotide comprising a sequence encoding a variant , fragment, or fusion of the native polypeptides of the instant invention.

BRIEF DESCRIPTION OF THE TABLES

The sequences of exemplary SDFs and polypeptides corresponding to the coding sequences of the instant invention are described in Table 1 and Table 2. Table 1 refers to a number of Maximum Length Sequences" or MLS." Each MLS corresponds to the longest cDNA obtained, either by cloning or by the prediction from genomic sequence. The sequence of the MLS is the cDNA sequence as described in the Av subsection of Table 1.

Table 1 includes the following information relating to each MLS:

I. cDNA Sequence

A. 5' UTR

B. Coding Sequence

C. 3' UTR

<u>I. cDNA SEQUENCE</u>

Table 1 indicates which sequence in Table 2 represents the sequence of each MLS. The MLS sequence can comprise 5' and 3' UTR as well as coding sequences. In addition, specific cDNA clone numbers also are included in Table 1 when the MLS sequence relates to a specific cDNA clone.

A. 5' UTR

The location of the 5' UTR can be determined by comparing the most 5' MLS sequence with the corresponding genomic sequence as indicated in Table 1. The sequence that matches, beginning at any of the transcriptional start sites and ending at the last nucleotide before any of the translational start sites corresponds to the 5' UTR.

B. Coding Region

The coding region is the sequence in any open reading frame found in the MLS. Coding regions of interest are indicated in the PolyP SEQ subsection Table 1.

C. 3' UTR

The location of the 3' UTR can be determined by comparing the most 3' MLS sequence with the corresponding genomic sequence as indicated in Table 1. The sequence that matches, beginning at the translational stop site and ending at the last nucleotide of the MLS corresponds to the 3' UTR.

II. GENOMIC SEQUENCE

Further, Table 1 indicates the specific gi" number of the genomic sequence if the sequence resides in a public databank. For each genomic sequence, Table 1 indicates which regions are included in the MLS. These regions can include the 5' and 3' UTRs as well as the coding sequence of the MLS. See, for example, the scheme below:

```
                 Region 1                 Region 2          Region 3
         ______________________         __________       ______________
---------| 5' UTR    | Exon |---------| Exon  |--------| Exon | 3' UTR |-------
         ______________________         __________       ______________
    ^                              ^                  ^                          ^
    |                  ^           |                  |          ^               |
Promoter               |         Intron             Intron       |
               Translational                                 Stop Codon
                Start Site
```

Table 1 reports the first and last base of each region that are included in an MLS sequence. An example is shown below:

gi No. 47000:

37102 ... 37497

37593 ... 37925

The numbers indicate that the MLS contains the following sequences from two regions of gi No. 47000; a first region including bases 37102-37497, and a second region including bases 37593-37925.

A. EXON SEQUENCES

The location of the exons can be determined by comparing the sequence of the regions from the genomic sequences with the corresponding MLS sequence as indicated by Table 1.

i. INITIAL EXON

To determine the location of the initial exon, information from the

(1) polypeptide sequence section;

(2) cDNA polynucleotide section; and

(3) the genomic sequence section

of Table 1 is used. First, the polypeptide section will indicate where the translational start site is located in the MLS sequence. The MLS sequence can be matched to the genomic sequence that corresponds to the MLS. Based on the match between the MLS and corresponding genomic sequences , the location of the translational start site can be determined in one of the regions of the genomic sequence. The location of this translational start site is the start of the first exon.

Generally, the last base of the exon of the corresponding genomic region, in which the translational start site was located, will represent the end of the initial exon. In some cases, the initial exon will end with a stop codon, when the initial exon is the only exon.

In the case when sequences representing the MLS are in the positive strand of the corresponding genomic sequence, the last base will be a larger number than the first base. When the sequences representing the MLS are in the negative strand of the corresponding genomic sequence, then the last base will be a smaller number than the first base.

ii. INTERNAL EXONS

Except for the regions that comprise the 5' and 3' UTRs, initial exon, and terminal exon, the remaining genomic regions that match the MLS sequence are the internal exons. Specifically, the bases defining the boundaries of the remaining regions also define the intron/exon junctions of the internal exons.

iii. TERMINAL EXON

As with the initial exon, the location of the terminal exon is determined with information from the

(1) polypeptide sequence section;

(2) cDNA polynucleotide section; and

(3) the genomic sequence section

of Table 1. The polypeptide section will indicate where the stop codon is located in the MLS sequence. The MLS sequence can be matched to the corresponding genomic sequence. Based on the match between MLS and corresponding genomic sequences, the location of the

stop codon can be determined in one of the regions of the genomic sequence. The location of this stop codon is the end of the terminal exon. Generally, the first base of the exon of the corresponding genomic region that matches the cDNA sequence, in which the stop codon was located, will represent the beginning of the terminal exon. In some cases, the translational start site will represent the start of the terminal exon, which will be the only exon.

In the case when the MLS sequences are in the positive strand of the corresponding genomic sequence, the last base will be a larger number than the first base. When the MLS sequences are in the negative strand of the corresponding genomic sequence, then the last base will be a smaller number than the first base.

B. INTRON SEQUENCES

In addition, the introns corresponding to the MLS are defined by identifying the genomic sequence located between the regions where the genomic sequence comprises exons. Thus, introns are defined as starting one base downstream of a genomic region comprising an exon, and end one base upstream from a genomic region comprising an exon.

C. PROMOTER SEQUENCES

As indicated below, promoter sequences corresponding to the MLS are defined as sequences upstream of the first exon; more usually, as sequences upstream of the first of multiple transcription start sites; even more usually as sequences about 2,000 nucleotides upstream of the first of multiple transcription start sites.

III. LINK of cDNA SEQUENCES to CLONE IDs

As noted above, Table 1 identifies the cDNA clone(s) that relate to each MLS. The MLS sequence can be longer than the sequences included in the cDNA clones. In such a case, Table 1 indicates the region of the MLS that is included in the clone. If either the 5' or 3' termini of the cDNA clone sequence is the same as the MLS sequence, no mention will be made.

IV. Multiple Transcription Start Sites

Initiation of transcription can occur at a number of sites of the gene. Table 1 indicates the possible multiple transcription sites for each gene. In Table 1, the location of the transcription start sites can be either a positive or negative number.

The positions indicated by positive numbers refer to the transcription start sites as located in the MLS sequence. The negative numbers indicate the transcription start site within the genomic sequence that corresponds to the MLS.

To determine the location of the transcription start sites with the negative numbers, the MLS sequence is aligned with the corresponding genomic sequence. In the instances when a public genomic sequence is referenced, the relevant corresponding genomic sequence can be found by direct reference to the nucleotide sequence indicated by the gi" number shown in the public genomic DNA section of Table 1. When the position is a negative number, the transcription start site is located in the corresponding genomic sequence upstream of the base that matches the beginning of the MLS sequence in the alignment. The negative number is relative to the first base of the MLS sequence which matches the genomic sequence corresponding to the relevant gi" number.

In the instances when no public genomic DNA is referenced, the relevant nucleotide sequence for alignment is the nucleotide sequence associated with the amino acid sequence designated by gi" number of the later PolyP SEQ subsection.

V. Polypeptide Sequences

The PolyP SEQ subsection lists SEQ ID NOs and Ceres SEQ ID NO for polypeptide sequences corresponding to the coding sequence of the MLS sequence and the location of the translational start site with the coding sequence of the MLS sequence.

The MLS sequence can have multiple translational start sites and can be capable of producing more than one polypeptide sequence.

A. Signal Peptide

Table 1 also indicates in subsection (B) the cleavage site of the putative signal peptide of the polypeptide corresponding to the coding sequence of the MLS sequence. Typically, signal peptide coding sequences comprise a sequence encoding the first residue of the polypeptide to the cleavage site residue.

B. Domains

Subsection (C) provides information regarding identified domains (where present) within the polypeptide and (where present) a name for the polypeptide domain.

C. Related Polypeptides

Subsection (Dp) provides (where present) information concerning amino acid sequences that are found to be related and have some percentage of sequence identity to the polypeptide sequences of Table 1 and Table 2. These related sequences are identified by a gi" number.

VI. Related Polynucleotide Sequences

Subsection (Dn) provides polynucleotide sequences (where present) that are related to and have some percentage of sequence identity to the MLS or corresponding genomic sequence.

Abbreviation	Description
Max Len. Seq.	Maximum Length Sequence
rel to	Related to
Clone Ids	Clone ID numbers
Pub gDNA	Public Genomic DNA
gi No.	gi number
Gen. seq. in cDNA	Genomic Sequence in cDNA (Each region for a single gene prediction is listed on a separate line. In the case of multiple gene predictions, the group of regions relating to a single prediction are separated by a blank line)
(Ac) cDNA SEQ	cDNA sequence
- Pat. Appln. SEQ ID NO	Patent Application SEQ ID NO:
- Ceres SEQ ID NO: 1673877	Ceres SEQ ID NO:
- SEQ # w. TSS	Location within the cDNA sequence, SEQ ID NO:, of Transcription Start Sites which are listed below
- Clone ID #: # -> #	Clone ID comprises bases # to # of the cDNA Sequence
PolyP SEQ	Polypeptide Sequence
- Pat. Appln. SEQ ID NO:	Patent Application SEQ ID NO:
- Ceres SEQ ID NO	Ceres SEQ ID NO:
- Loc. SEQ ID NO: @ nt.	Location of translational start site in cDNA of SEQ ID NO: at nucleotide number
(C) Pred. PP Nom. & Annot.	Nomination and Annotation of Domains within Predicted Polypeptide(s)
- (Title)	Name of Domain
- Loc. SEQ ID NO #: # -> # aa.	Location of the domain within the polypeptide of SEQ ID NO: from # to # amino acid residues.
(Dp) Rel. AA SEQ	Related Amino Acid Sequences
- Align. NO	Alignment number
- gi No	Gi number

Abbreviation	Description
- Desp.	Description
- % Idnt.	Percent identity
- Align. Len.	Alignment Length
- Loc. SEQ ID NO: # -> # aa	Location within SEQ ID NO: from # to # amino acid residue.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to (I) polynucleotides and methods of use thereof, such as

IA. Probes, Primers and Substrates;

IB. Methods of Detection and Isolation;

- B.1. Hybridization;
- B.2. Methods of Mapping;
- B.3. Southern Blotting;
- B.4. Isolating cDNA from Related Organisms;
- B.5. Isolating and/or Identifying Orthologous Genes

IC. Methods of Inhibiting Gene Expression

- C.1. Antisense
- C.2. Ribozyme Constructs;
- C.3. Chimeraplasts;
- C.4 Co-Suppression;
- C.5. Transcriptional Silencing
- C.6. Other Methods to Inhibit Gene Expression

ID. Methods of Functional Analysis;

IE. Promoter Sequences and Their Use;

IF. UTRs and/or Intron Sequences and Their Use; and

IG. Coding Sequences and Their Use.

The invention also relates to (II) polypeptides and proteins and methods of use thereof, such as IIA. Native Polypeptides and Proteins

- A.1 Antibodies
- A.2 In Vitro Applications

IIB. Polypeptide Variants, Fragments and Fusions

B.1 Variants

B.2 Fragments

B.3 Fusions

The invention also includes (III) methods of modulating polypeptide production, such as

IIIA. Suppression

A.1 Antisense

A.2 Ribozymes

A.3 Co-suppression

A.4 Insertion of Sequences into the Gene to be Modulated

A.5 Promoter Modulation

A.6 Expression of Genes containing Dominant-Negative Mutations

IIIB. Enhanced Expression

B.1 Insertion of an Exogenous Gene

B.2 Promoter Modulation

The invention further concerns (IV) gene constructs and vector construction, such as

IVA. Coding Sequences

IVB. Promoters

IVC. Signal Peptides

The invention still further relates to

V Transformation Techniques

Definitions

Allelic variant An allelic variant" is an alternative form of the same SDF, which resides at the same chromosomal locus in the organism. Allelic variations can occur in any portion of the gene sequence, including regulatory regions. Allelic variants can arise by normal genetic variation in a population. Allelic variants can also be produced by genetic

engineering methods. An allelic variant can be one that is found in a naturally occurring plant, including a cultivar or ecotype. An allelic variant may or may not give rise to a phenotypic change, and may or may not be expressed. An allele can result in a detectable change in the phenotype of the trait represented by the locus. A phenotypically silent allele can give rise to a product.

Alternatively spliced messages Within the context of the current invention, alternatively spliced messages" refers to mature mRNAs originating from a single gene with variations in the number and/or identity of exons, introns and/or intron-exon junctions.

Chimeric The term chimeric" is used to describe genes, as defined supra, or contructs wherein at least two of the elements of the gene or construct, such as the promoter and the coding sequence and/or other regulatory sequences and/or filler sequences and/or complements thereof, are heterologous to each other.

Constitutive Promoter: Promoters referred to herein as "constitutive promoters" actively promote transcription under most, but not necessarily all, environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcript initiation region and the 1' or 2' promoter derived from T-DNA of *Agrobacterium tumefaciens,* and other transcription initiation regions from various plant genes, such as the maize ubiquitin-1 promoter, known to those of skill.

Coordinately Expressed: The term coordinately expressed," as used in the current invention, refers to genes that are expressed at the same or a similar time and/or stage and/or under the same or similar environmental conditions.

Domain: Domains are fingerprints or signatures that can be used to characterize protein families and/or parts of proteins. Such fingerprints or signatures can comprise conserved (1) primary sequence, (2) secondary structure, and/or (3) three-dimensional conformation. Generally, each domain has been associated with either a family of proteins or motifs. Typically, these families and/or motifs have been correlated with specific *in-vitro* and/or *in-vivo* activities. A domain can be any length, including the entirety of the sequence of a protein. Detailed descriptions of the domains, associated families and motifs, and correlated activities of the polypeptides of the instant invention are described below.

Usually, the polypeptides with designated domain(s) can exhibit at least one activity that is exhibited by any polypeptide that comprises the same domain(s).

Endogenous The term endogenous," within the context of the current invention refers to any polynucleotide, polypeptide or protein sequence which is a natural part of a cell or organisms regenerated from said cell.

Exogenous Exogenous," as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation (of dicots - *e.g.* Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983); of monocots, representative papers are those by Escudero et al., *Plant J.* 10:355 (1996), Ishida et al., *Nature Biotechnology* 14:745 (1996), May et al., *Bio/Technology* 13:486 (1995)), biolistic methods (Armaleo et al., *Current Genetics* 17:97 1990)), electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T_0 for the primary transgenic plant and T_1 for the first generation. The term exogenous" as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

Filler sequence: As used herein, filler sequence" refers to any nucleotide sequence that is inserted into DNA construct to evoke a particular spacing between particular components such as a promoter and a coding region and may provide an additional attribute such as a restriction enzyme site.

Gene: The term gene," as used in the context of the current invention, encompasses all regulatory and coding sequence contiguously associated with a single hereditary unit with a genetic function (see SCHEMATIC 1). Genes can include non-coding sequences that modulate the genetic function that include, but are not limited to, those that specify polyadenylation, transcriptional regulation, DNA conformation, chromatin conformation, extent and position of base methylation and binding sites of proteins that control all of these. Genes comprised of exons" (coding sequences), which may be interrupted by introns"

(non-coding sequences), encode proteins. A gene's genetic function may require only RNA expression or protein production, or may only require binding of proteins and/or nucleic acids without associated expression. In certain cases, genes adjacent to one another may share sequence in such a way that one gene will overlap the other. A gene can be found within the genome of an organism, artificial chromosome, plasmid, vector, etc., or as a separate isolated entity.

Gene Family: Gene family" is used in the current invention to describe a group of functionally related genes, each of which encodes a separate protein.

Heterologous sequences: Heterologous sequences" are those that are not operatively linked or are not contiguous to each other in nature. For example, a promoter from corn is considered heterologous to an *Arabidopsis* coding region sequence. Also, a promoter from a gene encoding a growth factor from corn is considered heterologous to a sequence encoding the corn receptor for the growth factor. Regulatory element sequences, such as UTRs or 3' end termination sequences that do not originate in nature from the same gene as the coding sequence originates from, are considered heterologous to said coding sequence. Elements operatively linked in nature and contiguous to each other are not heterologous to each other. On the other hand, these same elements remain operatively linked but become heterologous if other filler sequence is placed between them. Thus, the promoter and coding sequences of a corn gene expressing an amino acid transporter are not heterologous to each other, but the promoter and coding sequence of a corn gene operatively linked in a novel manner are heterologous.

Homologous gene In the current invention, homologous gene" refers to a gene that shares sequence similarity with the gene of interest. This similarity may be in only a fragment of the sequence and often represents a functional domain such as, examples including without limitation a DNA binding domain, a domain with tyrosine kinase activity, or the like. The functional activities of homologous genes are not necessarily the same.

Inducible Promoter An inducible promoter" in the context of the current invention refers to a promoter which is regulated under certain conditions, such as light, chemical concentration, protein concentration, conditions in an organism, cell, or organelle, etc. A typical example of an inducible promoter, which can be utilized with the polynucleotides of the present invention, is PARSK1, the promoter from the *Arabidopsis* gene encoding a serine-threonine

kinase enzyme, and which promoter is induced by dehydration, abscissic acid and sodium chloride (Wang and Goodman, *Plant J.* 8:37 (1995)) Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

Intergenic region Intergenic region," as used in the current invention, refers to nucleotide sequence occurring in the genome that separates adjacent genes.

Mutant gene In the current invention, mutant" refers to a heritable change in DNA sequence at a specific location. Mutants of the current invention may or may not have an associated identifiable function when the mutant gene is transcribed.

Orthologous Gene In the current invention orthologous gene" refers to a second gene that encodes a gene product that performs a similar function as the product of a first gene. The orthologous gene may also have a degree of sequence similarity to the first gene. The orthologous gene may encode a polypeptide that exhibits a degree of sequence similarity to a polypeptide corresponding to a first gene. The sequence similarity can be found within a functional domain or along the entire length of the coding sequence of the genes and/or their corresponding polypeptides.

Percentage of sequence identity "Percentage of sequence identity," as used herein, is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of the polynucleotide or amino acid sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), by

computerized implementations of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection. Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. The term "substantial sequence identity" between polynucleotide or polypeptide sequences refers to polynucleotide or polypeptide comprising a sequence that has at least 80% sequence identity, preferably at least 85%, more preferably at least 90% and most preferably at least 95%, even more preferably, at least 96%, 97%, 98% or 99% sequence identity compared to a reference sequence using the programs.

Plant Promoter A plant promoter" is a promoter capable of initiating transcription in plant cells and can drive or facilitate transcription of a fragment of the SDF of the instant invention or a coding sequence of the SDF of the instant invention. Such promoters need not be of plant origin. For example, promoters derived from plant viruses, such as the CaMV35S promoter or from *Agrobacterium tumefaciens* such as the T-DNA promoters, can be plant promoters. A typical example of a plant promoter of plant origin is the maize ubiquitin-1 (ubi-1)promoter known to those of skill.

Promoter: The term "promoter," as used herein, refers to a region of sequence determinants located upstream from the start of transcription of a gene and which are involved in recognition and binding of RNA polymerase and other proteins to initiate and modulate transcription. A basal promoter is the minimal sequence necessary for assembly of a transcription complex required for transcription initiation. Basal promoters frequently include a TATA box" element usually located between 15 and 35 nucleotides upstream from the site of initiation of transcription. Basal promoters also sometimes include a CCAAT box" element (typically a sequence CCAAT) and/or a GGGCG sequence, usually located between 40 and 200 nucleotides, preferably 60 to 120 nucleotides, upstream from the start site of transcription.

Public sequence: The term public sequence ," as used in the context of the instant application, refers to any sequence that has been deposited in a publicly accessible database. This term encompasses both amino acid and nucleotide sequences. Such sequences are

publicly accessible, for example, on the BLAST databases on the NCBI FTP web site (accessible at ncbi.nlm.gov/blast). The database at the NCBI GTP site utilizes gi" numbers assigned by NCBI as a unique identifier for each sequence in the databases, thereby providing a non-redundant database for sequence from various databases, including GenBank, EMBL, DBBJ, (DNA Database of Japan) and PDB (Brookhaven Protein Data Bank).

Regulatory Sequence The term regulatory sequence," as used in the current invention, refers to any nucleotide sequence that influences transcription or translation initiation and rate, and stability and/or mobility of the transcript or polypeptide product. Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start site, termination sequence, polyadenylation sequence, introns, certain sequences within a coding sequence, etc.

Related Sequences: Related sequences" refer to either a polypeptide or a nucleotide sequence that exhibits some degree of sequence similarity with a sequence described by Table 1 and Table 2.

Scaffold Attachment Region (SAR) As used herein, scaffold attachment region" is a DNA sequence that anchors chromatin to the nuclear matrix or scaffold to generate loop domains that can have either a transcriptionally active or inactive structure (Spiker and Thompson (1996) Plant Physiol. 110: 15-21).

Sequence-determined DNA fragments (SDFs) Sequence-determined DNA fragments" as used in the current invention are isolated sequences of genes, fragments of genes, intergenic regions or contiguous DNA from plant genomic DNA or cDNA or RNA the sequence of which has been determined.

Signal Peptide A signal peptide" as used in the current invention is an amino acid sequence that targets the protein for secretion, for transport to an intracellular compartment or organelle or for incorporation into a membrane. Signal peptides are indicated in the tables and a more detailed description located below.

Specific Promoter In the context of the current invention, specific promoters" refers to a subset of inducible promoters that have a high preference for being induced in a specific tissue or cell and/or at a specific time during development of an organism. By high preference" is meant at least 3-fold, preferably 5-fold, more preferably at least 10-fold still more preferably at least 20-fold, 50-fold or 100-fold increase in transcription in the desired tissue over the transcription in any other tissue. Typical examples of temporal and/or tissue specific promoters of plant origin that can be used with the polynucleotides of the present invention, are: PTA29, a promoter which is capable of driving gene transcription specifically in tapetum and only during anther development (Koltonow et al., *Plant Cell* 2:1201 (1990); RCc2 and RCc3, promoters that direct root-specific gene transcription in rice (Xu et al., *Plant Mol. Biol.* 27:237 (1995); TobRB27, a root-specific promoter from tobacco (Yamamoto et al., *Plant Cell* 3:371 (1991)). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues or organs, such as root, ovule, fruit, seeds, or flowers. Other suitable promoters include those from genes encoding storage proteins or the lipid body membrane protein, oleosin. A few root-specific promoters are noted above.

Stringency "Stringency" as used herein is a function of probe length, probe composition (G + C content), and salt concentration, organic solvent concentration, and temperature of hybridization or wash conditions. Stringency is typically compared by the parameter T_m, which is the temperature at which 50% of the complementary molecules in the hybridization are hybridized, in terms of a temperature differential from T_m. High stringency conditions are those providing a condition of T_m - 5°C to T_m - 10°C. Medium or moderate stringency conditions are those providing T_m - 20°C to T_m - 29°C. Low stringency conditions are those providing a condition of T_m - 40°C to T_m - 48°C. The relationship of hybridization conditions to T_m (in °C) is expressed in the mathematical equation

$$T_m = 81.5 - 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N) \quad (1)$$

where N is the length of the probe. This equation works well for probes 14 to 70 nucleotides in length that are identical to the target sequence. The equation below for T_m of DNA-DNA hybrids is useful for probes in the range of 50 to greater than 500 nucleotides, and for conditions that include an organic solvent (formamide).

$$T_m = 81.5 + 16.6 \log \{[Na^+]/(1+0.7[Na^+])\} + 0.41(\%G+C) - 500/L\ 0.63(\%formamide) \quad (2)$$

where L is the length of the probe in the hybrid. (P. Tijessen, Hybridization with Nucleic Acid Probes" in Laboratory Techniques in Biochemistry and Molecular Biology, P.C. vand der Vliet, ed., c. 1993 by Elsevier, Amsterdam.) The T_m of equation (2) is affected by the nature of the hybrid; for DNA-RNA hybrids T_m is 10-15°C higher than calculated, for RNA-RNA hybrids T_m is 20-25°C higher. Because the T_m decreases about 1 °C for each 1% decrease in homology when a long probe is used (Bonner et al., *J. Mol. Biol.* 81:123 (1973)), stringency conditions can be adjusted to favor detection of identical genes or related family members.

Equation (2) is derived assuming equilibrium and therefore, hybridizations according to the present invention are most preferably performed under conditions of probe excess and for sufficient time to achieve equilibrium. The time required to reach equilibrium can be shortened by inclusion of a hybridization accelerator such as dextran sulfate or another high volume polymer in the hybridization buffer.

Stringency can be controlled during the hybridization reaction or after hybridization has occurred by altering the salt and temperature conditions of the wash solutions used. The formulas shown above are equally valid when used to compute the stringency of a wash solution. Preferred wash solution stringencies lie within the ranges stated above; high stringency is 5-8°C below T_m, medium or moderate stringency is 26-29°C below T_m and low stringency is 45-48°C below T_m.

Substantially free of A composition containing A is substantially free of B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight. For example, a plant gene or DNA sequence can be considered substantially free of other plant genes or DNA sequences.

Translational start site In the context of the current invention, a translational start site" is usually an ATG in the cDNA transcript, more usually the first ATG. A single cDNA, however, may have multiple translational start sites.

Transcription start site Transcription start site" is used in the current invention to describe the point at which transcription is initiated. This point is typically located about 25 nucleotides downstream from a TFIID binding site, such as a TATA box. Transcription can

initiate at one or more sites within the gene, and a single gene may have multiple transcriptional start sites, some of which may be specific for transcription in a particular cell-type or tissue.

Untranslated region (UTR) A UTR" is any contiguous series of nucleotide bases that is transcribed, but is not translated. These untranslated regions may be associated with particular functions such as increasing mRNA message stability. Examples of UTRs include, but are not limited to polyadenylation signals, terminations sequences, sequences located between the transcriptional start site and the first exon (5' UTR) and sequences located between the last exon and the end of the mRNA (3' UTR).

Variant: The term variant" is used herein to denote a polypeptide or protein or polynucleotide molecule that differs from others of its kind in some way. For example, polypeptide and protein variants can consist of changes in amino acid sequence and/or charge and/or post-translational modifications (such as glycosylation, etc).

DETAILED DESCRIPTION OF THE INVENTION

I. Polynucleotides

Exemplified SDFs of the invention represent fragments of the genome of corn, wheat, rice, soybean or *Arabidopsis* and/or represent mRNA expressed from that genome. The isolated nucleic acid of the invention also encompasses corresponding fragments of the genome and/or cDNA complement of other organisms as described in detail below.

Polynucleotides of the invention can be isolated from polynucleotide libraries using primers comprising sequence similar to those described by Table 1 and Table 2. See, for example, the methods described in Sambrook et al., supra.

Alternatively, the polynucleotides of the invention can be produced by chemical synthesis. Such synthesis methods are described below.

It is contemplated that the nucleotide sequences presented herein may contain some small percentage of errors. These errors may arise in the normal course of determination of nucleotide sequences. Sequence errors can be corrected by obtaining seeds deposited under the accession numbers cited above, propagating them, isolating genomic DNA or appropriate mRNA from the resulting plants or seeds thereof, amplifying the relevant fragment of the genomic DNA or mRNA using primers having a sequence that flanks the erroneous sequence, and sequencing the amplification product.

I.A. Probes, Primers and Substrates

SDFs of the invention can be applied to substrates for use in array applications such as, but not limited to, assays of global gene expression, for example under varying conditions of development, growth conditions. The arrays can also be used in diagnostic or forensic methods (WO95/35505, US 5,445,943 and US 5,410,270).

Probes and primers of the instant invention will hybridize to a polynucleotide comprising a sequence in Tables 1 and 2. Though many different nucleotide sequences can encode an amino acid sequence, the sequences of Tables 1 and 2 are generally preferred for encoding polypeptides of the invention. However, the sequence of the probes and/or primers of the instant invention need not be identical to those in Tables 1 and 2 or the complements thereof. For example, some variation in probe or primer sequence and/or length can allow additional family members to be detected, as well as orthologous genes and more taxonomically distant related sequences. Similarly, probes and/or primers of the invention can include additional nucleotides that serve as a label for detecting the formed duplex or for subsequent cloning purposes.

Probe length will vary depending on the application. For use as primers, probes are 12-40 nucleotides, preferably 18-30 nucleotides long. For use in mapping, probes are preferably 50 to 500 nucleotides, preferably 100-250 nucleotides long. For Southern hybridizations, probes as long as several kilobases can be used as explained below.

The probes and/or primers can be produced by synthetic procedures such as the triester method of Matteucci et al. *J. Am. Chem. Soc.* 103:3185(1981); or according to Urdea et al. *Proc. Natl. Acad.* 80:7461 (1981) or using commercially available automated oligonucleotide synthesizers.

I.B. Methods of Detection and Isolation

The polynucleotides of the invention can be utilized in a number of methods known to those skilled in the art as probes and/or primers to isolate and detect polynucleotides, including, without limitation: Southerns, Northerns, Branched DNA hybridization assays, polymerase chain reaction, and microarray assays, and variations thereof. Specific methods given by way of examples, and discussed below include:

Hybridization

Methods of Mapping

Southern Blotting

Isolating cDNA from Related Organisms

Isolating and/or Identifying Orthologous Genes.

Also, the nucleic acid molecules of the invention can used in other methods, such as high density oligonucleotide hybridizing assays, described, for example, in U.S. Pat. Nos. 6,004,753; 5,945,306; 5,945,287; 5,945,308; 5,919,686; 5,919,661; 5,919,627; 5,874,248; 5,871,973; 5,871,971; and 5,871,930; and PCT Pub. Nos. WO 9946380; WO 9933981; WO 9933870; WO 9931252; WO 9915658; WO 9906572; WO 9858052; WO 9958672; and WO 9810858.

B.1. Hybridization

The isolated SDFs of Tables 1 and 2 of the present invention can be used as probes and/or primers for detection and/or isolation of related polynucleotide sequences through hybridization. Hybridization of one nucleic acid to another constitutes a physical property that defines the subject SDF of the invention and the identified related sequences. Also, such hybridization imposes structural limitations on the pair. A good general discussion of the factors for determining hybridization conditions is provided by Sambrook et al. ("Molecular Cloning, a Laboratory Manual, 2nd ed., c. 1989 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; *see esp.,* chapters 11 and 12). Additional considerations and details of the physical chemistry of hybridization are provided by G.H. Keller and M.M. Manak DNA Probes", 2nd Ed. pp. 1-25, c. 1993 by Stockton Press, New York, NY.

Depending on the stringency of the conditions under which these probes and/or primers are used, polynucleotides exhibiting a wide range of similarity to those in Tables 1 and 2 can be detected or isolated. When the practitioner wishes to examine the result of membrane hybridizations under a variety of stringencies, an efficient way to do so is to perform the hybridization under a low stringency condition, then to wash the hybridization membrane under increasingly stringent conditions.

When using SDFs to identify orthologous genes in other species, the practitioner will preferably adjust the amount of target DNA of each species so that, as nearly as is practical, the same number of genome equivalents are present for each species examined. This prevents faint signals from species having large genomes, and thus small numbers of genome equivalents per mass of DNA, from erroneously being interpreted as absence of the corresponding gene in the genome.

The probes and/or primers of the instant invention can also be used to detect or isolate nucleotides that are identical" to the probes or primers. Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below.

Isolated polynucleotides within the scope of the invention also include allelic variants of the specific sequences presented in Tables 1 and 2. The probes and/or primers of the invention can also be used to detect and/or isolate polynucleotides exhibiting at least 80% sequence identity with the sequences of Tables 1 and 2 or fragments thereof.

With respect to nucleotide sequences, degeneracy of the genetic code provides the possibility to substitute at least one base of the base sequence of a gene with a different base without causing the amino acid sequence of the polypeptide produced from the gene to be changed. Hence, the DNA of the present invention may also have any base sequence that has been changed from a sequence in Tables 1 and 2 by substitution in accordance with degeneracy of genetic code. References describing codon usage include: Carels *et al.*, *J. Mol. Evol.* 46: 45 (1998) and Fennoy *et al.*, *Nucl. Acids Res.* 21(23): 5294 (1993).

B.2. Mapping

The isolated SDF DNA of the invention can be used to create various types of genetic and physical maps of the genome of corn, Arabidopsis, soybean, rice, wheat, or other plants. Some SDFs may be absolutely associated with particular phenotypic traits, allowing construction of gross genetic maps. While not all SDFs will immediately be associated with a phenotype, all SDFs can be used as probes for identifying polymorphisms associated with phenotypes of interest. Briefly, one method of mapping involves total DNA isolation from individuals. It is subsequently cleaved with one or more restriction enzymes, separated according to mass, transferred to a solid support, hybridized with SDF DNA and the pattern

of fragments compared. Polymorphisms associated with a particular SDF are visualized as differences in the size of fragments produced between individual DNA samples after digestion with a particular restriction enzyme and hybridization with the SDF. After identification of polymorphic SDF sequences, linkage studies can be conducted. By using the individuals showing polymorphisms as parents in crossing programs, F2 progeny recombinants or recombinant inbreds, for example, are then analyzed. The order of DNA polymorphisms along the chromosomes can be determined based on the frequency with which they are inherited together versus independently. The closer two polymorphisms are together in a chromosome the higher the probability that they are inherited together. Integration of the relative positions of all the polymorphisms and associated marker SDFs can produce a genetic map of the species, where the distances between markers reflect the recombination frequencies in that chromosome segment.

The use of recombinant inbred lines for such genetic mapping is described for *Arabidopsis* by Alonso-Blanco et al. (*Methods in Molecular Biology*, vol.82, *Arabidopsis Protocols*", pp. 137-146, J.M. Martinez-Zapater and J. Salinas, eds., c. 1998 by Humana Press, Totowa, NJ) and for corn by Burr (Mapping Genes with Recombinant Inbreds", pp. 249-254. *In* Freeling, M. and V. Walbot (Ed.), *The Maize Handbook*, c. 1994 by Springer-Verlag New York, Inc.: New York, NY, USA; Berlin Germany; Burr et al. *Genetics* (1998) 118: 519; Gardiner, J. et al., (1993) *Genetics* 134: 917). This procedure, however, is not limited to plants and can be used for other organisms (such as yeast) or for individual cells.

The SDFs of the present invention can also be used for simple sequence repeat (SSR) mapping. Rice SSR mapping is described by Morgante et al. (*The Plant Journal* (1993) 3: 165), Panaud et al. (*Genome* (1995) 38: 1170); Senior et al. (*Crop Science* (1996) 36: 1676), Taramino et al. (*Genome* (1996) 39: 277) and Ahn et al. (*Molecular and General Genetics* (1993) 241: 483-90). SSR mapping can be achieved using various methods. In one instance, polymorphisms are identified when sequence specific probes contained within an SDF flanking an SSR are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals of interest. Here, a change in the number of tandem repeats between the SSR-flanking sequences produces differently sized fragments (U.S. Patent 5,766,847). Alternatively, polymorphisms can be identified by using the PCR fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (U.H. Refseth et al., (1997) *Electrophoresis* 18: 1519).

Genetic and physical maps of crop species have many uses. For example, these maps can be used to devise positional cloning strategies for isolating novel genes from the mapped crop species. In addition, because the genomes of closely related species are largely syntenic (that is, they display the same ordering of genes within the genome), these maps can be used to isolate novel alleles from relatives of crop species by positional cloning strategies.

The various types of maps discussed above can be used with the SDFs of the invention to identify Quantitative Trait Loci (QTLs). Many important crop traits, such as the solids content of tomatoes, are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, oftentimes on different chromosomes, and generally exhibit multiple alleles at each locus. The SDFs of the invention can be used to identify QTLs and isolate specific alleles as described by de Vicente and Tanksley (*Genetics* 134:585 (1993)). In addition to isolating QTL alleles in present crop species, the SDFs of the invention can also be used to isolate alleles from the corresponding QTL of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once a desired allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs (for review *see* Tanksley and McCouch, *Science* 277:1063 (1997)).

In another embodiment, the SDFs can be used to help create physical maps of the genome of corn, *Arabidopsis* and related species. Where SDFs have been ordered on a genetic map, as described above, they can be used as probes to discover which clones in large libraries of plant DNA fragments in YACs, BACs, etc. contain the same SDF or similar sequences, thereby facilitating the assignment of the large DNA fragments to chromosomal positions. Subsequently, the large BACs, YACs, etc. can be ordered unambiguously by more detailed studies of their sequence composition (e.g. Marra et al. (1997) Genomic Research 7:1072-1084) and by using their end or other sequences to find the identical sequences in other cloned DNA fragments. The overlapping of DNA sequences in this way allows large contigs of plant sequences to be built that, when sufficiently extended, provide a complete physical map of a chromosome. Sometimes the SDFs themselves will provide the means of joining cloned sequences into a contig.

The patent publication WO95/35505 and U.S. Patents 5,445,943 and 5,410,270 describe scanning multiple alleles of a plurality of loci using hybridization to arrays of oligonucleotides. These techniques are useful for each of the types of mapping discussed above.

Following the procedures described above and using a plurality of the SDFs of the present invention, any individual can be genotyped. These individual genotypes can be used for the identification of particular cultivars, varieties, lines, ecotypes and genetically modified plants or can serve as tools for subsequent genetic studies involving multiple phenotypic traits.

B.3 Southern Blot Hybridization

The sequences from Tables 1 and 2 can be used as probes for various hybridization techniques. These techniques are useful for detecting target polynucleotides in a sample or for determining whether transgenic plants, seeds or host cells harbor a gene or sequence of interest and thus might be expected to exhibit a particular trait or phenotype.

In addition, the SDFs from the invention can be used to isolate additional members of gene families from the same or different species and/or orthologous genes from the same or different species. This is accomplished by hybridizing an SDF to, for example, a Southern blot containing the appropriate genomic DNA or cDNA. Given the resulting hybridization data, one of ordinary skill in the art could distinguish and isolate the correct DNA fragments by size, restriction sites, sequence and stated hybridization conditions from a gel or from a library.

Identification and isolation of orthologous genes from closely related species and alleles within a species is particularly desirable because of their potential for crop improvement. Many important crop traits, such as the solid content of tomatoes, result from the combined interactions of the products of several genes residing at different loci in the genome. Generally, alleles at each of these loci can make quantitative differences to the trait. By identifying and isolating numerous alleles for each locus from within or different species, transgenic plants with various combinations of alleles can be created and the effects of the combinations measured. Once a more favorable allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs (Tanksley et al. *Science* 277:1063(1997)).

The results from hybridizations of the SDFs of the invention to, for example, Southern blots containing DNA from another species can also be used to generate restriction fragment maps for the corresponding genomic regions. These maps provide additional information about the relative positions of restriction sites within fragments, further distinguishing mapped DNA from the remainder of the genome.

Physical maps can be made by digesting genomic DNA with different combinations of restriction enzymes.

Probes for Southern blotting to distinguish individual restriction fragments can range in size from 15 to 20 nucleotides to several thousand nucleotides. More preferably, the probe is 100 to 1,000 nucleotides long for identifying members of a gene family when it is found that repetitive sequences would complicate the hybridization. For identifying an entire corresponding gene in another species, the probe is more preferably the length of the gene, typically 2,000 to 10,000 nucleotides, but probes 50-1,000 nucleotides long might be used. Some genes, however, might require probes up to 1,500 nucleotides long or overlapping probes constituting the full-length sequence to span their lengths.

Also, while it is preferred that the probe be homogeneous with respect to its sequence, it is not necessary. For example, as described below, a probe representing members of a gene family having diverse sequences can be generated using PCR to amplify genomic DNA or RNA templates using primers derived from SDFs that include sequences that define the gene family.

For identifying corresponding genes in another species, the next most preferable probe is a cDNA spanning the entire coding sequence, which allows all of the mRNA-coding fragment of the gene to be identified. Probes for Southern blotting can easily be generated from SDFs by making primers having the sequence at the ends of the SDF and using corn or *Arabidopsis* genomic DNA as a template. In instances where the SDF includes sequence conserved among species, primers including the conserved sequence can be used for PCR with genomic DNA from a species of interest to obtain a probe.

Similarly, if the SDF includes a domain of interest, that fragment of the SDF can be used to make primers and, with appropriate template DNA, used to make a probe to identify genes containing the domain. Alternatively, the PCR products can be resolved, for example by gel electrophoresis, and cloned and/or sequenced. Using Southern hybridization, the variants of the domain among members of a gene family, both within and across species, can be examined.

B.4.1 Isolating DNA from Related Organisms

The SDFs of the invention can be used to isolate the corresponding DNA from other organisms. Either cDNA or genomic DNA can be isolated. For isolating genomic DNA, a lambda, cosmid, BAC or YAC, or other large insert genomic library from the plant of interest can be constructed using standard molecular biology techniques as described in detail by

Sambrook et al. 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York) and by Ausubel et al. 1992 (Current Protocols in Molecular Biology, Greene Publishing, New York).

To screen a phage library, for example, recombinant lambda clones are plated out on appropriate bacterial medium using an appropriate *E. coli* host strain. The resulting plaques are lifted from the plates using nylon or nitrocellulose filters. The plaque lifts are processed through denaturation, neutralization, and washing treatments following the standard protocols outlined by Ausubel et al. (1992). The plaque lifts are hybridized to either radioactively labeled or non-radioactively labeled SDF DNA at room temperature for about 16 hours, usually in the presence of 50% formamide and 5X SSC (sodium chloride and sodium citrate) buffer and blocking reagents. The plaque lifts are then washed at 42°C with 1% Sodium Dodecyl Sulfate (SDS) and at a particular concentration of SSC. The SSC concentration used is dependent upon the stringency at which hybridization occurred in the initial Southern blot analysis performed. For example, if a fragment hybridized under medium stringency (e.g., Tm - 20°C), then this condition is maintained or preferably adjusted to a less stringent condition (e.g., Tm-30°C) to wash the plaque lifts. Positive clones show detectable hybridization e.g., by exposure to X-ray films or chromogen formation. The positive clones are then subsequently isolated for purification using the same general protocol outlined above. Once the clone is purified, restriction analysis can be conducted to narrow the region corresponding to the gene of interest. The restriction analysis and succeeding subcloning steps can be done using procedures described by, for example Sambrook et al. (1989) cited above.

The procedures outlined for the lambda library are essentially similar to those used for YAC library screening, except that the YAC clones are harbored in bacterial colonies. The YAC clones are plated out at reasonable density on nitrocellulose or nylon filters supported by appropriate bacterial medium in petri plates. Following the growth of the bacterial clones, the filters are processed through the denaturation, neutralization, and washing steps following the procedures of Ausubel et al. 1992. The same hybridization procedures for lambda library screening are followed.

To isolate cDNA, similar procedures using appropriately modified vectors are employed. For instance, the library can be constructed in a lambda vector appropriate for cloning cDNA such as λgt11. Alternatively, the cDNA library can be made in a plasmid vector. cDNA for cloning can be prepared by any of the methods known in the art, but is

preferably prepared as described above. Preferably, a cDNA library will include a high proportion of full-length clones.

B. 5. Isolating and/or Identifying Orthologous Genes

Probes and primers of the invention can be used to identify and/or isolate polynucleotides related to those in Tables 1 and 2. Related polynucleotides are those that are native to other plant organisms and exhibit either similar sequence or encode polypeptides with similar biological activity. One specific example is an orthologous gene. Orthologous genes have the same functional activity. As such, orthologous genes may be distinguished from homologous genes. The percentage of identity is a function of evolutionary separation and, in closely related species, the percentage of identity can be 98 to 100%. The amino acid sequence of a protein encoded by an orthologous gene can be less than 75% identical, but tends to be at least75% or at least 80% identical, more preferably at least 90%, most preferably at least 95% identical to the amino acid sequence of the reference protein.

To find orthologous genes, the probes are hybridized to nucleic acids from a species of interest under low stringency conditions, preferably one where sequences containing as much as 40-45% mismatches will be able to hybridize. This condition is established by T_m - 40°C to Tm - 48°C (*see* below). Blots are then washed under conditions of increasing stringency. It is preferable that the wash stringency be such that sequences that are 85 to 100% identical will hybridize. More preferably, sequences 90 to 100% identical will hybridize and most preferably only sequences greater than 95% identical will hybridize. One of ordinary skill in the art will recognize that, due to degeneracy in the genetic code, amino acid sequences that are identical can be encoded by DNA sequences as little as 67% identical or less. Thus, it is preferable, for example, to make an overlapping series of shorter probes, on the order of 24 to 45 nucleotides, and individually hybridize them to the same arrayed library to avoid the problem of degeneracy introducing large numbers of mismatches.

As evolutionary divergence increases, genome sequences also tend to diverge. Thus, one of skill will recognize that searches for orthologous genes between more divergent species will require the use of lower stringency conditions compared to searches between closely related species. Also, degeneracy of the genetic code is more of a problem for searches in the genome of a species more distant evolutionarily from the species that is the source of the SDF probe sequences.

Therefore the method described in Bouckaert et al., U.S. Ser. No. 60/121,700 Atty. Dkt. No. 2750-117P, Client Dkt. No. 00010.001, filed February 25, 1999, hereby

incorporated in its entirety by reference, can be applied to the SDFs of the present invention to isolate related genes from plant species which do not hybridize to the corn *Arabidopsis*, soybean, rice, wheat, and other plant sequences of Tables 1 and 2.

Identification of the relationship of nucleotide or amino acid sequences among plant species can be done by comparing the nucleotide or amino acid sequences of SDFs of the present application with nucleotide or amino acid sequences of other SDFs such as those present in applications listed in the table below:

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0301P	80002.001	9/4/98	60/099,672
United States	2750-0300P	80001.001	9/4/98	60/099,671
United States	2750-0302P	80003.001	9/11/98	60/099,933
United States	2750-0304P	80004.001	9/17/98	60/100,864
United States	2750-0305P	80005.001	9/18/98	60/101,042
United States	2750-0306P	80006.001	9/21/98	60/101,255
United States	2750-0307P	80007.001	9/24/98	60/101,682
United States	2750-0308P	80008.001	9/30/98	60/102,533
United States	2750-0309P	80009.001	9/30/98	60/102,460
United States	2750-0311P	80011.001	10/5/98	60/103,141
United States	2750-0310P	80010.001	10/5/98	60/103,116
United States	2750-0312P	80012.001	10/6/98	60/103,215
United States	2750-0313P	80013.001	10/8/98	60/103,554
United States	2750-0314P	80014.001	10/9/98	60/103,574
United States	2750-0315P	80015.001	10/13/98	60/103,907
United States	2750-0316P	80016.001	10/14/98	60/104,268
United States	2750-0317P	80017.001	10/16/98	60/104,680
United States	2750-0318P	80018.001	10/19/98	60/104,828
United States	2750-0319P	80019.001	10/20/98	60/105,008
United States	2750-0320P	80020.001	10/21/98	60/105,142
United States	2750-0321P	80021.001	10/22/98	60/105,533
United States	2750-0322P	80022.001	10/26/98	60/105,571
United States	2750-0323P	80023.001	10/27/98	60/105,815
United States	2750-0324P	80024.001	10/29/98	60/106,105
United States	2750-0325P	80025.001	10/30/98	60/106,218
United States	2750-0326P	80026.001	11/2/98	60/106,685
United States	2750-0327P	80027.001	11/6/98	60/107,282
United States	2750-0329P	80029.001	11/9/98	60/107,719
United States	2750-0328P	80028.001	11/9/98	60/107,720
United States	2750-0330P	80030.001	11/10/98	60/107,836
United States	2750-0331P	80031.001	11/12/98	60/108,190
United States	2750-0332P	80032.001	11/16/98	60/108,526
United States	2750-0333P	80033.001	11/17/98	60/108,901
United States	2750-0335P	80035.001	11/19/98	60/109,127
United States	2750-0334P	80034.001	11/19/98	60/109,124
United States	2750-0336P	80036.001	11/20/98	60/109,267
United States	2750-0337P	80037.001	11/23/98	60/109,594
United States	2750-0338P	80038.001	11/25/98	60/110,053
United States	2750-0339P	80039.001	11/25/98	60/110,050

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0340P	80040.001	11/27/98	60/110,158
United States	2750-0341P	80041.001	11/30/98	60/110,263
United States	2750-0342P	80042.001	12/1/98	60/110,495
United States	2750-0343P	80043.001	12/2/98	60/110,626
United States	2750-0344P	80044.001	12/3/98	60/110,701
United States	2750-0345P	80045.001	12/7/98	60/111,339
United States	2750-0346P	80046.001	12/9/98	60/111,589
United States	2750-0347P	80047.001	12/10/98	60/111,782
United States	2750-0348P	80048.001	12/11/98	60/111,812
United States	2750-0349P	80049.001	12/14/98	60/112,096
United States	2750-0350P	80050.001	12/15/98	60/112,224
United States	2750-0351P	80051.001	12/16/98	60/112,624
United States	2750-0352P	80052.001	12/17/98	60/112,862
United States	2750-0353P	80053.001	12/18/98	60/112,912
United States	2750-0354P	80054.001	12/21/98	60/113,248
United States	2750-0355P	80055.001	12/22/98	60/113,522
United States	2750-0356P	80056.001	12/23/98	60/113,826
United States	2750-0357P	80057.001	12/28/98	60/113,998
United States	2750-0358P	80058.001	12/29/98	60/114,384
United States	2750-0359P	80059.001	12/30/98	60/114,455
United States	2750-0360P	80060.001	1/4/99	60/114,740
United States	2750-0361P	80061.001	1/6/99	60/114,866
United States	2750-0364P	80064.001	1/7/99	60/115,151
United States	2750-0365P	80065.001	1/7/99	60/115,155
United States	2750-0362P	80062.001	1/7/99	60/115,153
United States	2750-0363P	80063.001	1/7/99	60/115,152
United States	2750-0367P	80067.001	1/7/99	60/115,154
United States	2750-0366P	80066.001	1/7/99	60/115,156
United States	2750-0370P	80070.001	1/8/99	60/115,293
United States	2750-0369P	80069.001	1/8/99	60/115,365
United States	2750-0368P	80068.001	1/8/99	60/115,364
United States	2750-0371P	80071.001	1/11/99	60/115,339
United States	2750-0372P	80072.001	1/12/99	60/115,518
United States	2750-0373P	80073.001	1/13/99	60/115,847
United States	2750-0374P	80074.001	1/14/99	60/115,905
United States	2750-0375P	80075.001	1/15/99	60/116,383
United States	2750-0376P	80076.001	1/15/99	60/116,384
United States	2750-0378P	80078.001	1/19/99	60/116,340
United States	2750-0377P	80077.001	1/19/99	60/116,329
United States	2750-0380P	80080.001	1/21/99	60/116,672
United States	2750-0379P	80079.001	1/21/99	60/116,674
United States	2750-0381P	80081.001	1/22/99	60/116,960
United States	2750-0382P	80082.001	1/22/99	60/116,962
United States	2750-0383P	80083.001	1/28/99	60/117,756
United States	2750-0384P	80084.001	2/3/99	60/118,672
United States	2750-0385P	80085.001	2/4/99	60/118,808
United States	2750-0386P	80086.001	2/5/99	60/118,778
United States	2750-0387P	80087.001	2/8/99	60/119,029
United States	2750-0388P	80088.001	2/9/99	60/119,332
United States	2750-0389P	80089.001	2/10/99	60/119,462

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0391P	80091.001	2/12/99	60/119,922
United States	2750-0393P	80093.001	2/16/99	60/120,198
United States	2750-0392P	80092.001	2/16/99	60/120,196
United States	2750-0394P	80094.001	2/18/99	60/120,583
United States	2750-0395P	80095.001	2/22/99	60/121,072
United States	2750-0396P	80096.001	2/23/99	60/121,334
United States	2750-0397P	80097.001	2/24/99	60/121,470
United States	2750-0390P	80090.001	2/25/99	60/121,825
United States	2750-0398P	80098.001	2/25/99	60/121,704
United States	2750-0399P	80099.001	2/26/99	60/122,107
United States	2750-0400P	80100.001	3/1/99	60/122,266
United States	2750-0401P	80101.001	3/2/99	60/122,568
United States	2750-0402P	80102.001	3/3/99	60/122,611
United States	2750-0403P	80103.001	3/4/99	60/121,775
United States	2750-0405P	80105.001	3/5/99	60/123,180
United States	2750-0404P	80104.001	3/5/99	60/123,534
United States	2750-0406P	80106.001	3/9/99	60/123,680
United States	2750-0407P	80107.001	3/9/99	60/123,548
United States	2750-0408P	80108.001	3/10/99	60/123,715
United States	2750-0409P	80109.001	3/10/99	60/123,726
United States	2750-0410P	80110.001	3/11/99	60/124,263
United States	2750-0411P	80111.001	3/12/99	60/123,941
United States	2750-0412P	80112.001	3/23/99	60/125,788
United States	2750-0413P	80113.001	3/25/99	60/126,264
United States	2750-0414P	80114.001	3/29/99	60/126,785
United States	2750-0415P	80115.001	4/1/99	60/127,462
United States	2750-0416P	91000.001	4/6/99	60/128,234
United States	2750-0417P	91001.001	4/8/99	60/128,714
United States	2750-0418P	80118.001	4/16/99	60/129,845
United States	2750-0420P	80120.001	4/19/99	60/130,077
United States	2750-0421P	80121.001	4/21/99	60/130,449
United States	2750-0303P	80115.002	4/23/99	60/130,510
United States	2750-0422P	80122.001	4/23/99	60/130,891
United States	2750-0423P	80123.001	4/28/99	60/131,449
United States	2750-0424P	80124.001	4/30/99	60/132,407
United States	2750-0425P	80125.001	4/30/99	60/132,048
United States	2750-0426P	80126.001	5/4/99	60/132,484
United States	2750-0427P	80127.001	5/5/99	60/132,485
United States	2750-0428P	91002.001	5/6/99	60/132,487
United States	2750-0429P	80129.001	5/6/99	60/132,486
United States	2750-0430P	80130.001	5/7/99	60/132,863
United States	2750-0431P	80131.001	5/11/99	60/134,256
United States	2750-0435P	80117.001	5/14/99	60/134,218
United States	2750-0433P	00025.001	5/14/99	60/134,221
United States	2750-0432P	91006.001	5/14/99	60/134,370
United States	2750-0434P	80116.001	5/14/99	60/134,219
United States	2750-0436P	91007.001	5/18/99	60/134,768
United States	2750-0437P	91008.001	5/19/99	60/134,941
United States	2750-0438P	91009.001	5/20/99	60/135,124
United States	2750-0439P	91010.001	5/21/99	60/135,353

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0440P	91011.001	5/24/99	60/135,629
United States	2750-0441P	91012.001	5/25/99	60/136,021
United States	2750-0442P	91013.001	5/27/99	60/136,392
United States	2750-0444P	91014.001	5/28/99	60/136,782
United States	2750-0445P	91015.001	6/1/99	60/137,222
United States	2750-0446P	91016.001	6/3/99	60/137,528
United States	2750-0447P	91017.001	6/4/99	60/137,502
United States	2750-0449P	91018.001	6/7/99	60/137,724
United States	2750-0450P	91019.001	6/8/99	60/138,094
United States	2750-0457P	00033.001	6/10/99	60/138,540
United States	2750-0458P	00033.002	6/10/99	60/138,847
United States	2750-0463P	00034.001	6/14/99	60/139,119
United States	2750-0462P	80132.012	6/16/99	60/139,452
United States	2750-0461P	80132.011	6/16/99	60/139,453
United States	2750-0464P	00037.001	6/17/99	60/139,492
United States	2750-0465P	00038.001	6/18/99	60/139,763
United States	2750-0460P	80132.010	6/18/99	60/139,455
United States	2750-0459P	80132.009	6/18/99	60/139,463
United States	2750-0466P	00039.001	6/18/99	60/139,750
United States	2750-0455P	80132.007	6/18/99	60/139,460
United States	2750-0454P	80132.006	6/18/99	60/139,457
United States	2750-0453P	80132.005	6/18/99	60/139,462
United States	2750-0452P	80132.004	6/18/99	60/139,461
United States	2750-0451P	80132.003	6/18/99	60/139,459
United States	2750-0443P	80132.001	6/18/99	60/139,458
United States	2750-0448P	80132.002	6/18/99	60/139,454
United States	2750-0456P	80132.008	6/18/99	60/139,456
United States	2750-0467P	00042.001	6/21/99	60/139,817
United States	2750-0468P	00043.001	6/22/99	60/139,899
United States	2750-0469P	00044.001	6/23/99	60/140,354
United States	2750-0470P	00042.002	6/23/99	60/140,353
United States	2750-0471P	00045.001	6/24/99	60/140,695
United States	2750-0472P	00046.001	6/28/99	60/140,823
United States	2750-0473P	00048.001	6/29/99	60/140,991
United States	2750-0474P	00049.001	6/30/99	60/141,287
United States	2750-0476P	00051.001	7/1/99	60/142,154
United States	2750-0475P	00050.001	7/1/99	60/141,842
United States	2750-0477P	00052.001	7/2/99	60/142,055
United States	2750-0478P	00053.001	7/6/99	60/142,390
United States	2750-0479P	00054.001	7/8/99	60/142,803
United States	2750-0480P	00058.001	7/9/99	60/142,920
United States	2750-0481P	00059.001	7/12/99	60/142,977
United States	2750-0482P	00060.001	7/13/99	60/143,542
United States	2750-0489P	00061.001	7/14/99	60/143,624
United States	2750-0490P	00062.001	7/15/99	60/144,005
United States	2750-0485P	80134.003	7/16/99	60/144,086
United States	2750-0486P	80134.004	7/16/99	60/144,085
United States	2750-0488P	80134.006	7/19/99	60/144,332
United States	2750-0495P	80134.013	7/19/99	60/144,335
United States	2750-0497P	00064.001	7/19/99	60/144,325

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0492P	80134.008	7/19/99	60/144,331
United States	2750-0496P	80134.014	7/19/99	60/144,334
United States	2750-0494P	80134.010	7/19/99	60/144,333
United States	2750-0499P	80134.012	7/20/99	60/144,352
United States	2750-0502P	80135.002	7/20/99	60/144,884
United States	2750-0500P	00065.001	7/20/99	60/144,632
United States	2750-0503P	00066.001	7/21/99	60/144,814
United States	2750-0483P	80134.001	7/21/99	60/145,088
United States	2750-0484P	80134.002	7/21/99	60/145,086
United States	2750-0493P	80134.009	7/22/99	60/145,087
United States	2750-0504P	00067.001	7/22/99	60/145,192
United States	2750-0487P	80134.005	7/22/99	60/145,089
United States	2750-0491P	80134.007	7/22/99	60/145,085
United States	2750-0501P	80135.001	7/23/99	60/145,224
United States	2750-0498P	80134.011	7/23/99	60/145,145
United States	2750-0505P	00069.001	7/23/99	60/145,218
United States	2750-0506P	00070.001	7/26/99	60/145,276
United States	2750-0507P	80136.001	7/27/99	60/145,918
United States	2750-0508P	80136.002	7/27/99	60/145,919
United States	2750-0509P	00071.001	7/27/99	60/145,913
United States	2750-0510P	00072.001	7/28/99	60/145,951
United States	2750-0513P	00073.001	8/2/99	60/146,386
United States	2750-0512P	80137.002	8/2/99	60/146,389
United States	2750-0511P	80137.001	8/2/99	60/146,388
United States	2750-0514P	00074.001	8/3/99	60/147,038
United States	2750-0517P	80138.002	8/4/99	60/147,302
United States	2750-0515P	00076.001	8/4/99	60/147,204
United States	2750-0519P	80136.003	8/5/99	60/147,192
United States	2750-0518P	00077.001	8/5/99	60/147,260
United States	2750-0520P	00079.001	8/6/99	60/147,416
United States	2750-0516P	80138.001	8/6/99	60/147,303
United States	2750-0521P	00080.001	8/9/99	60/147,493
United States	2750-0523P	80139.002	8/9/99	60/147,935
United States	2750-0522P	80139.001	8/10/99	60/148,171
United States	2750-0524P	00081.001	8/11/99	60/148,319
United States	2750-0530P	00082.001	8/12/99	60/148,341
United States	2750-0526P	80141.002	8/12/99	60/148,342
United States	2750-0527P	80141.003	8/12/99	60/148,340
United States	2750-0528P	80141.004	8/12/99	60/148,337
United States	2750-0525P	80141.001	8/12/99	60/148,347
United States	2750-0529P	00083.001	8/13/99	60/148,565
United States	2750-0532P	80142.002	8/13/99	60/148,684
United States	2750-0531P	80142.001	8/16/99	60/149,368
United States	2750-0533P	80001.002	8/17/99	60/149,927
United States	2750-0534P	80001.003	8/17/99	60/149,928
United States	2750-0535P	80001.004	8/17/99	60/149,926
United States	2750-0536P	80001.005	8/17/99	60/149,925
United States	2750-0537P	00084.001	8/17/99	60/149,175
United States	2750-0538P	00085.001	8/18/99	60/149,426
United States	2750-0542P	00087.001	8/20/99	60/149,723

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0541P	80143.002	8/20/99	60/149,929
United States	2750-0539P	00086.001	8/20/99	60/149,722
United States	2750-0543P	00088.001	8/23/99	60/149,902
United States	2750-0540P	80143.001	8/23/99	60/149,930
United States	2750-0544P	00089.001	8/25/99	60/150,566
United States	2750-0547P	00090.001	8/26/99	60/150,884
United States	2750-0546P	80144.002	8/27/99	60/151,066
United States	2750-0545P	80144.001	8/27/99	60/151,065
United States	2750-0548P	00091.001	8/27/99	60/151,080
United States	2750-0549P	00092.001	8/30/99	60/151,303
United States	2750-0552P	00093.001	8/31/99	60/151,438
United States	2750-0553P	00094.001	9/1/99	60/151,930
International	2750-0551F(PC)	80001.100	9/3/99	99/204,38
United States	2750-0550P	80001.006	9/3/99	09/391,631
United States	2750-0554P	00095.001	9/7/99	60/152,363
United States	2750-0555P	00096.001	9/10/99	60/153,070
United States	2750-0556P	00098.001	9/13/99	60/153,758
United States	2750-0557P	00099.001	9/15/99	60/154,018
United States	2750-0558P	00101.001	9/16/99	60/154,039
United States	2750-0559P	00102.001	9/20/99	60/154,779
United States	2750-0560P	00103.001	9/22/99	60/155,139
United States	2750-0561P	00104.001	9/23/99	60/155,486
United States	2750-0562P	00105.001	9/24/99	60/155,659
United States	2750-0563P	00106.001	9/28/99	60/156,458
United States	2750-0564P	00107.001	9/29/99	60/156,596
United States	2750-0570P	00108.001	10/4/99	60/157,117
United States	2750-0571P	00109.001	10/5/99	60/157,753
International	2750-0569F(PC)	80010.102	10/5/99	99/228,53
International	2750-0568F(PC)	80010.101	10/5/99	99/228,54
International	2750-0567F(PC)	80010.100	10/5/99	99/228,55
United States	2750-0566P	80010.003	10/5/99	09/412,922
United States	2750-0565P	80010.002	10/5/99	09/413,198
United States	2750-0572P	00110.001	10/6/99	60/157,865
United States	2750-0575P	00111.001	10/7/99	60/158,029
United States	2750-0576P	00112.001	10/8/99	60/158,232
United States	2750-0577P	00113.001	10/12/99	60/158,369
United States	2750-0574P	80145.002	10/13/99	60/159,295
United States	2750-0579P	80146.002	10/13/99	60/159,293
United States	2750-0583P	80148.002	10/13/99	60/159,294
United States	2750-0578P	80146.001	10/14/99	60/159,331
United States	2750-0573P	80145.001	10/14/99	60/159,330
United States	2750-0580P	80147.001	10/14/99	60/159,638
United States	2750-0581P	80147.002	10/14/99	60/159,637
United States	2750-0582P	80148.001	10/14/99	60/159,329
United States	2750-0584P	00116.001	10/18/99	60/159,584
United States	2750-0588P	00119.001	10/21/99	60/160,741
United States	2750-0589P	80150.001	10/21/99	60/160,768
United States	2750-0587P	80149.002	10/21/99	60/160,770
United States	2750-0586P	80149.001	10/21/99	60/160,814
United States	2750-0585P	00118.001	10/21/99	60/160,815

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0590P	80150.002	10/21/99	60/160,767
United States	2750-0591P	00120.001	10/22/99	60/160,980
United States	2750-0592P	80151.001	10/22/99	60/160,989
United States	2750-0593P	80151.002	10/22/99	60/160,981
United States	2750-0596P	80152.002	10/25/99	60/161,404
United States	2750-0594P	00121.001	10/25/99	60/161,405
United States	2750-0595P	80152.001	10/25/99	60/161,406
United States	2750-0599P	80153.002	10/26/99	60/161,359
United States	2750-0598P	80153.001	10/26/99	60/161,360
United States	2750-0597P	00122.001	10/26/99	60/161,361
United States	2750-0601P	00123.001	10/28/99	60/161,920
United States	2750-0602P	80154.001	10/28/99	60/161,992
United States	2750-0600P	80026.002	10/28/99	09/428,944
United States	2750-0603P	80154.002	10/28/99	60/161,993
United States	2750-0606P	80155.002	10/29/99	60/162,228
United States	2750-0605P	80155.001	10/29/99	60/162,142
United States	2750-0604P	00124.001	10/29/99	60/162,143
United States	2750-0607P	00125.001	11/1/99	60/162,894
United States	2750-0608P	80156.001	11/1/99	60/162,891
United States	2750-0609P	80156.002	11/1/99	60/162,895
United States	2750-0610P	00126.001	11/2/99	60/163,093
United States	2750-0611P	80157.001	11/2/99	60/163,092
United States	2750-0612P	80157.002	11/2/99	60/163,091
United States	2750-0613P	00127.001	11/3/99	60/163,249
United States	2750-0615P	80158.002	11/3/99	60/163,281
United States	2750-0614P	80158.001	11/3/99	60/163,248
United States	2750-0618P	80159.002	11/4/99	60/163,380
United States	2750-0616P	00128.001	11/4/99	60/163,379
United States	2750-0617P	80159.001	11/4/99	60/163,381
United States	2750-0619P	00129.001	11/8/99	60/164,146
United States	2750-0620P	80160.001	11/8/99	60/164,151
United States	2750-0621P	80160.002	11/8/99	60/164,150
United States	2750-0623P	80161.002	11/9/99	60/164,260
United States	2750-0625P	80162.002	11/9/99	60/164,259
United States	2750-0626P	80163.001	11/10/99	60/164,321
United States	2750-0629P	80164.001	11/10/99	60/164,545
United States	2750-0628P	00131.001	11/10/99	60/164,544
United States	2750-0627P	80163.002	11/10/99	60/164,318
United States	2750-0622P	80161.001	11/10/99	60/164,319
United States	2750-0630P	80164.002	11/10/99	60/164,548
United States	2750-0624P	80162.001	11/10/99	60/164,317
United States	2750-0633P	80165.002	11/12/99	60/164,960
United States	2750-0636P	80166.002	11/12/99	60/164,962
United States	2750-0634P	00133.001	11/12/99	60/164,870
United States	2750-0632P	80165.001	11/12/99	60/164,871
United States	2750-0631P	00132.001	11/12/99	60/164,961
United States	2750-0635P	80166.001	11/12/99	60/164,959
United States	2750-0637P	00134.001	11/15/99	60/164,927
United States	2750-0638P	80167.001	11/15/99	60/164,929
United States	2750-0639P	80167.002	11/15/99	60/164,926

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0641P	80168.001	11/16/99	60/165,671
United States	2750-0640P	00135.001	11/16/99	60/165,669
United States	2750-0642P	80168.002	11/16/99	60/165,661
United States	2750-0645P	80169.002	11/17/99	60/165,911
United States	2750-0643P	00136.001	11/17/99	60/165,919
United States	2750-0644P	80169.001	11/17/99	60/165,918
United States	2750-0646P	00137.001	11/18/99	60/166,157
United States	2750-0647P	80170.001	11/18/99	60/166,173
United States	2750-0648P	80170.002	11/18/99	60/166,158
United States	2750-0649P	00139.001	11/19/99	60/166,419
United States	2750-0650P	80171.001	11/19/99	60/166,411
United States	2750-0651P	80171.002	11/19/99	60/166,412
United States	2750-0652P	00140.001	11/22/99	60/166,733
United States	2750-0653P	80172.001	11/22/99	60/166,750
United States	2750-0655P	80173.002	11/23/99	60/167,362
United States	2750-0656P	00141.001	11/24/99	60/167,233
United States	2750-0657P	80174.001	11/24/99	60/167,234
United States	2750-0654P	80173.001	11/24/99	60/167,382
United States	2750-0658P	80174.002	11/24/99	60/167,235
United States	2750-0659P	00142.001	11/30/99	60/167,904
United States	2750-0660P	80175.001	11/30/99	60/167,908
United States	2750-0661P	80175.002	11/30/99	60/167,902
United States	2750-0663P	00143.001	12/1/99	60/168,232
United States	2750-0662P	80042.002	12/1/99	09/451,320
United States	2750-0665P	80176.002	12/1/99	60/168,231
United States	2750-0664P	80176.001	12/1/99	60/168,233
United States	2750-0667P	80177.001	12/2/99	60/168,549
United States	2750-0668P	80177.002	12/2/99	60/168,548
United States	2750-0666P	00144.001	12/2/99	60/168,546
United States	2750-0669P	00145.001	12/3/99	60/168,675
United States	2750-0670P	80178.001	12/3/99	60/168,673
United States	2750-0671P	80178.002	12/3/99	60/168,674
United States	2750-0674P	80179.002	12/7/99	60/169,302
United States	2750-0672P	00147.001	12/7/99	60/169,298
United States	2750-0673P	80179.001	12/7/99	60/169,278
United States	2750-0676P	80180.002	12/8/99	60/169,691
United States	2750-0675P	80180.001	12/8/99	60/169,692
United States	2750-0677P	00149.001	12/16/99	60/171,107
United States	2750-0678P	80181.001	12/16/99	60/171,114
United States	2750-0679P	80181.002	12/16/99	60/171,098
United States	2750-0683P	80060.002	1/4/00	09/478,081
International	2750-0686F(PC)	80070.100	1/7/00	00/004,66
United States	2750-0684P	80070.002	1/7/00	09/479,221
United States	2750-0685P	80183.002	1/19/00	60/176,867
United States	2750-0688P	80184.002	1/19/00	60/176,910
United States	2750-0681P	80182.002	1/19/00	60/176,866
United States	2750-0689P	00152.001	1/26/00	60/178,166
United States	2750-0691P	80185.001	1/27/00	60/177,666
United States	2750-0690P	00153.001	1/27/00	60/178,547
United States	2750-0682P	80183.001	1/27/00	60/178,546

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0680P	80182.001	1/27/00	60/178,544
United States	2750-0687P	80184.001	1/27/00	60/178,545
United States	2750-0692P	00155.001	1/28/00	60/178,754
United States	2750-0693P	80186.001	1/28/00	60/178,755
United States	2750-0695P	00157.001	2/1/00	60/179,395
United States	2750-0696P	80187.001	2/1/00	60/179,388
United States	2750-0694P	80084.002	2/3/00	09/497,191
United States	2750-0697P	00158.001	2/3/00	60/180,039
United States	2750-0698P	80188.001	2/3/00	60/180,139
United States	2750-0699P	00159.001	2/4/00	60/180,206
United States	2750-0700P	80189.001	2/4/00	60/180,207
United States	2750-0701P	00160.001	2/7/00	60/180,695
United States	2750-0702P	80190.001	2/7/00	60/180,696
United States	2750-0703P	00161.001	2/9/00	60/181,228
United States	2750-0704P	80191.001	2/9/00	60/181,214
United States	2750-0705P	00162.001	2/10/00	60/181,476
United States	2750-0706P	80192.001	2/10/00	60/181,551
United States	2750-0707P	00163.001	2/15/00	60/182,477
United States	2750-0708P	80193.001	2/15/00	60/182,516
United States	2750-0712P	00164.001	2/15/00	60/182,512
United States	2750-0713P	80194.001	2/15/00	60/182,478
United States	2750-0714P	00165.001	2/17/00	60/183,166
United States	2750-0715P	80195.001	2/17/00	60/183,165
United States	2750-0716P	00167.001	2/24/00	60/184,667
United States	2750-0717P	80196.001	2/24/00	60/184,658
Canada	2750-0709F(CA)	80090.102	2/25/00	23/006,92
United States	2750-0719P	00168.001	2/25/00	60/185,118
United States	2750-0718P	91022.001	2/25/00	60/185,140
United States	2750-0720P	80197.001	2/25/00	60/185,119
Europe	2750-0709F(EP)	80090.103	2/25/00	00/301,439
Mexico	2750-0709F(MX)	80090.101	2/25/00	00/001,973
United States	2750-0709P	80090.002	2/25/00	09/513,996
United States	2750-0722P	00169.001	2/28/00	60/185,396
United States	2750-0721P	91023.001	2/28/00	60/185,398
United States	2750-0723P	80198.001	2/28/00	60/185,397
United States	2750-0724P	91024.001	2/29/00	60/185,750
United States	2750-0725P	00170.001	3/1/00	
United States	2750-0726P	80199.001	3/1/00	60/186,296
United States	2750-0727P	91025.001	3/1/00	60/186,277
United States	2750-0710P	80100.002	3/1/00	09/517,537
United States	2750-0730P	80201.001	3/2/00	60/186,387
United States	2750-0711P	00171.001	3/2/00	60/186,390
United States	2750-0728P	80200.001	3/2/00	60/187,178
United States	2750-0729P	00172.001	3/2/00	60/186,386
United States	2750-0733P	80202.001	3/3/00	60/186,669
United States	2750-0731P	91026.001	3/3/00	60/186,670
United States	2750-0732P	00173.001	3/3/00	60/186,748
United States	2750-0734P	00174.001	3/7/00	60/187,378
United States	2750-0735P	91027.001	3/7/00	60/187,379
United States	2750-0736P	00175.001	3/8/00	60/187,896

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0737P	80203.001	3/8/00	60/187,888
United States	2750-0738P	91028.001	3/9/00	60/187,985
United States	2750-0741P	91030.001	3/10/00	60/188,174
United States	2750-0742P	00178.001	3/10/00	60/188,185
United States	2750-0743P	80205.001	3/10/00	60/188,175
United States	2750-0739P	00177.001	3/10/00	60/188,187
United States	2750-0740P	80204.001	3/10/00	60/188,186
United States	2750-0744P	91031.001	3/13/00	60/188,687
United States	2750-0745P	00179.001	3/14/00	60/189,080
United States	2750-0746P	80206.001	3/14/00	60/189,052
United States	2750-0747P	91032.001	3/15/00	60/189,460
United States	2750-0748P	00180.001	3/15/00	60/189,461
United States	2750-0749P	80207.001	3/15/00	60/189,462
United States	2750-0755P	00181.001	3/16/00	60/189,953
United States	2750-0757P	91034.001	3/16/00	60/189,965
United States	2750-0754P	91033.001	3/16/00	60/189,958
United States	2750-0753P	80211.001	3/16/00	60/190,121
United States	2750-0752P	80210.001	3/16/00	60/189,948
United States	2750-0751P	80209.001	3/16/00	60/189,947
United States	2750-0750P	80208.001	3/16/00	60/190,120
United States	2750-0756P	80212.001	3/16/00	60/189,959
United States	2750-0762P	80214.001	3/20/00	60/190,089
United States	2750-0761P	00183.001	3/20/00	60/190,545
United States	2750-0760P	91035.001	3/20/00	60/190,060
United States	2750-0758P	00182.001	3/20/00	60/190,069
United States	2750-0759P	80213.001	3/20/00	60/190,070
United States	2750-0763P	00184.001	3/22/00	60/191,084
United States	2750-0764P	80215.001	3/22/00	60/191,097
United States	2750-0765P	91036.001	3/23/00	60/191,549
United States	2750-0766P	00185.001	3/23/00	60/191,543
United States	2750-0767P	80216.001	3/23/00	60/191,545
United States	2750-0769P	00186.001	3/24/00	60/191,823
United States	2750-0770P	80217.001	3/24/00	60/191,825
United States	2750-0768P	91037.001	3/24/00	60/191,826
United States	2750-0773P	80218.001	3/27/00	60/192,308
United States	2750-0772P	00187.001	3/27/00	60/192,421
United States	2750-0771P	91038.001	3/27/00	60/192,420
United States	2750-0775P	00188.001	3/29/00	60/192,940
United States	2750-0774P	91039.001	3/29/00	60/192,855
United States	2750-0776P	80219.001	3/29/00	60/192,941
United States	2750-0779P	80220.001	3/30/00	60/193,245
United States	2750-0777P	91040.001	3/30/00	60/193,243
United States	2750-0778P	00189.001	3/30/00	60/193,244
United States	2750-0780P	91041.001	3/31/00	60/193,469
United States	2750-0781P	00190.001	3/31/00	60/193,453
United States	2750-0782P	80221.001	3/31/00	60/193,455
United States	2750-0785P	91042.001	4/4/00	60/194,385
United States	2750-0786P	00191.001	4/4/00	60/194,404
United States	2750-0787P	80222.001	4/4/00	60/194,398
United States	2750-0789P	91043.001	4/5/00	60/194,682

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0792P	91044.001	4/5/00	60/194,698
United States	2750-0791P	80223.001	4/5/00	60/194,697
United States	2750-0790P	00192.001	4/5/00	60/194,683
Canada	2750-0783F(CA)	91000.102	4/6/00	02/302,828
Europe	2750-0783F(EP)	91000.101	4/6/00	00/302,919
Mexico	2750-0783F(MX)	91000.100	4/6/00	00/003,391
United States	2750-0793P	00193.001	4/6/00	60/194,874
United States	2750-0794P	80224.001	4/6/00	60/194,872
United States	2750-0796P	80225.001	4/6/00	60/195,045
United States	2750-0783P	91000.002	4/6/00	09/543,680
United States	2750-0784P	91045.001	4/6/00	60/194,884
United States	2750-0795P	00194.001	4/6/00	60/194,885
United States	2750-0798P	00195.001	4/7/00	60/195,283
United States	2750-0797P	91046.001	4/7/00	60/195,258
United States	2750-0799P	80226.001	4/7/00	60/195,257
United States	2750-0802P	91047.001	4/11/00	60/196,168
United States	2750-0801P	80227.002	4/11/00	60/196,211
United States	2750-0804P	80228.001	4/11/00	60/196,089
United States	2750-0803P	00196.001	4/11/00	60/196,169
United States	2750-0809P	80230.001	4/12/00	60/196,486
United States	2750-0808P	00200.001	4/12/00	60/196,485
United States	2750-0807P	80229.001	4/12/00	60/196,289
United States	2750-0806P	00197.001	4/12/00	60/196,487
United States	2750-0805P	91048.001	4/12/00	60/196,483
United States	2750-0800P	80227.001	4/12/00	60/196,212
United States	2750-0811P	80231.002	4/13/00	60/196,213
United States	2750-0814P	91049.001	4/14/00	60/197,397
United States	2750-0810P	80231.001	4/14/00	
United States	2750-0819P	80234.001	4/17/00	60/197,671
United States	2750-0812P	80232.001	4/17/00	60/197,870
United States	2750-0813P	80232.002	4/17/00	60/197,871
United States	2750-0815P	00201.001	4/17/00	60/197,687
United States	2750-0817P	91050.001	4/17/00	60/198,268
United States	2750-0816P	80233.001	4/17/00	60/197,678
United States	2750-0818P	00202.001	4/17/00	60/198,133
United States	2750-0822P	80235.001	4/19/00	60/198,373
United States	2750-0820P	91051.001	4/19/00	60/198,400
United States	2750-0821P	00203.001	4/19/00	60/198,386
United States	2750-0823P	91052.001	4/20/00	60/198,629
United States	2750-0824P	00204.001	4/20/00	60/198,619
United States	2750-0825P	80236.001	4/20/00	60/198,623
United States	2750-0828P	80237.001	4/21/00	60/198,763
United States	2750-0826P	91053.001	4/21/00	60/198,765
United States	2750-0827P	00206.001	4/21/00	60/198,767
United States	2750-0831P	80238.001	4/24/00	60/199,122
United States	2750-0830P	00207.001	4/24/00	60/199,124
United States	2750-0829P	91054.001	4/24/00	60/199,123
United States	2750-0832P	92001.001	4/26/00	60/200,034
United States	2750-0834P	00208.001	4/26/00	60/199,828
United States	2750-0833P	92002.001	4/26/00	60/200,031

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0835P	80239.001	4/26/00	60/199,818
United States	2750-0837P	80240.001	4/27/00	60/200,102
United States	2750-0836P	00210.001	4/27/00	60/200,103
United States	2750-0788P	80123.002	4/28/00	09/559,232
United States	2750-0846P	80243.002	4/28/00	60/200,773
United States	2750-0844P	80242.002	4/28/00	60/200,373
United States	2750-0845P	80243.001	5/1/00	60/201,016
United States	2750-0840P	91055.001	5/1/00	60/200,885
United States	2750-0839P	80241.001	5/1/00	60/200,879
United States	2750-0843P	80242.001	5/1/00	60/200,763
United States	2750-0848P	80244.002	5/1/00	60/200,761
United States	2750-0847P	80244.001	5/1/00	60/200,762
United States	2750-0842P	92002.002	5/1/00	60/201,018
United States	2750-0841P	92001.002	5/1/00	60/201,017
United States	2750-0849P	91056.001	5/2/00	60/201,279
United States	2750-0838P	00211.001	5/2/00	60/201,275
United States	2750-0850P	80245.001	5/2/00	60/201,305
United States	2750-0856P	91057.001	5/4/00	60/201,751
United States	2750-0857P	00212.001	5/4/00	60/201,740
Mexico	2750-0851F(MX)	91002.102	5/4/00	00/004,406
United States	2750-0858P	80246.001	5/4/00	60/201,750
United States	2750-0852P	80126.002	5/4/00	09/566,262
United States	2750-0851P	91002.002	5/5/00	09/565,308
Europe	2750-0851F(EP)	91002.101	5/5/00	00/303,770
Canada	2750-0851F(CA)	91002.100	5/5/00	02/305,695
United States	2750-0861P	80247.001	5/5/00	60/202,180
United States	2750-0855P	80130.002	5/5/00	09/565,310
United States	2750-0854P	80129.002	5/5/00	09/565,307
United States	2750-0853P	80127.002	5/5/00	09/565,309
United States	2750-0860P	00213.001	5/5/00	60/202,112
United States	2750-0859P	91058.001	5/5/00	60/202,178
United States	2750-0862P	00214.001	5/9/00	60/202,914
United States	2750-0864P	91059.001	5/9/00	60/202,915
United States	2750-0865P	00215.001	5/9/00	60/202,919
United States	2750-0866P	80249.001	5/9/00	60/202,634
United States	2750-0863P	80248.001	5/9/00	60/202,636
United States	2750-0878P	00216.001	5/10/00	60/202,968
United States	2750-0877P	91060.001	5/10/00	60/202,969
United States	2750-0879P	80252.001	5/10/00	60/202,963
United States	2750-0869P	80251.001	5/11/00	60/203,669
United States	2750-0871P	80131.002	5/11/00	09/572,408
United States	2750-0881P	00217.001	5/11/00	60/203,457
United States	2750-0880P	91061.001	5/11/00	60/203,458
United States	2750-0868P	80250.002	5/11/00	60/203,672
United States	2750-0867P	80250.001	5/11/00	60/203,671
United States	2750-0870P	80251.002	5/11/00	60/203,622
United States	2750-0882P	80253.001	5/11/00	60/203,279
Europe	2750-0875F(EP)	91006.101	5/12/00	00/304,017
Mexico	2750-0875F(MX)	91006.102	5/12/00	00/004,628
Canada	2750-0875F(CA)	91006.100	5/12/00	02/306,232

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0874P	80116.002	5/12/00	09/570,738
United States	2750-0873P	00025.002	5/12/00	09/570,582
United States	2750-0872P	80117.002	5/12/00	09/570,768
United States	2750-0883P	91062.001	5/12/00	60/203,911
United States	2750-0885P	80254.001	5/12/00	60/203,915
United States	2750-0884P	00219.001	5/12/00	60/203,916
United States	2750-0875P	91006.002	5/12/00	09/570,581
United States	2750-0887P	00220.001	5/15/00	60/204,388
United States	2750-0888P	80255.001	5/15/00	60/204,122
United States	2750-0886P	91063.001	5/15/00	60/204,395
United States	2750-0891P	00221.001	5/16/00	60/204,568
United States	2750-0892P	80256.001	5/16/00	60/204,569
United States	2750-0890P	92002.003	5/17/00	60/205,325
United States	2750-0893P	00222.001	5/17/00	60/204,830
United States	2750-0889P	92001.003	5/17/00	60/205,233
United States	2750-0894P	80257.001	5/17/00	60/204,829
United States	2750-0896P	80258.001	5/18/00	60/205,058
Canada	2750-0876F(CA)	91007.100	5/18/00	02/306,202
Europe	2750-0876F(EP)	91007.101	5/18/00	00/304,161
United States	2750-0895P	00223.001	5/18/00	60/205,201
United States	2750-0876P	91007.002	5/18/00	09/573,655
Mexico	2750-0876F(MX)	91007.102	5/18/00	00/004,850
United States	2750-0898P	80259.001	5/19/00	60/205,243
United States	2750-0897P	00224.001	5/19/00	60/205,242
United States	2750-0899P	91064.001	5/22/00	60/205,574
United States	2750-0900P	00225.001	5/22/00	60/205,572
United States	2750-0901P	80260.001	5/22/00	60/205,576
United States	2750-0903P	80261.001	5/23/00	60/206,319
United States	2750-0902P	00226.001	5/23/00	60/206,316
United States	2750-0904P	00227.001	5/24/00	60/206,553
United States	2750-0905P	80262.001	5/24/00	60/206,545
United States	2750-0906P	91065.001	5/25/00	60/206,988
United States	2750-0911P	80264.001	5/26/00	60/207,354
United States	2750-0910P	00229.001	5/26/00	60/207,239
United States	2750-0908P	80263.001	5/26/00	60/207,243
United States	2750-0909P	91066.001	5/26/00	60/207,242
United States	2750-0907P	00228.001	5/26/00	60/207,367
United States	2750-0914P	80265.001	5/30/00	60/207,329
United States	2750-0913P	00230.001	5/30/00	60/207,452
United States	2750-0912P	91067.001	5/30/00	60/207,291
United States	2750-0919P	91068.001	6/1/00	60/208,324
United States	2750-0921P	80268.001	6/1/00	60/208,312
United States	2750-0916P	80266.002	6/1/00	60/208,421
United States	2750-0918P	80267.002	6/1/00	60/208,648
United States	2750-0920P	00231.001	6/1/00	60/208,329
United States	2750-0917P	80267.001	6/2/00	60/208,649
United States	2750-0915P	80266.001	6/2/00	60/209,338
United States	2750-0926P	00233.001	6/5/00	60/208,921
United States	2750-0927P	80270.001	6/5/00	60/208,920
United States	2750-0925P	91070.001	6/5/00	60/208,917

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0924P	80269.001	6/5/00	60/208,918
United States	2750-0923P	00232.001	6/5/00	60/208,910
United States	2750-0922P	91069.001	6/5/00	60/208,919
United States	2750-0931P	80271.001	6/8/00	60/210,006
United States	2750-0930P	00234.001	6/8/00	60/210,012
United States	2750-0929P	91071.001	6/8/00	60/210,008
Mexico	2750-1037F(MX)	91000.103(MX)	6/9/00	00/005,741
United States	2750-0932P	00235.001	6/9/00	60/210,670
United States	2750-0928P	00033.003	6/9/00	09/592,459
Mexico	2750-0928F(MX)	00033.102	6/9/00	00/005,740
Canada	2750-0928F(CA)	00033.100	6/9/00	
United States	2750-0933P	80272.001	6/9/00	60/210,564
Europe	2750-0928F(EP)	00033.101	6/12/00	00/304,943
United States	2750-0936P	80273.001	6/13/00	60/211,214
United States	2750-0935P	00237.001	6/13/00	60/211,213
United States	2750-0937P	91072.001	6/13/00	60/211,210
Canada	2750-0934F(CA)	00034.100	6/14/00	
Europe	2750-0934F(EP)	00034.101	6/14/00	00/305,026
United States	2750-0934P	00034.002	6/14/00	09/593,710
Mexico	2750-0934F(MX)	00034.102	6/14/00	00/005,842
United States	2750-0938P	00238.001	6/15/00	60/211,539
United States	2750-0939P	80274.001	6/15/00	60/211,540
United States	2750-0940P	91074.001	6/15/00	60/211,538
Europe	2750-0942F(EP)	00038.101	6/16/00	
Europe	2750-0941F(EP)	00037.101	6/16/00	00/305,144
Mexico	2750-0941F(MX)	00037.102	6/16/00	00/005,950
Canada	2750-0943F(CA)	00039.100	6/16/00	
Europe	2750-0943F(EP)	00039.101	6/16/00	
United States	2750-0955P	80132.024	6/16/00	09/595,331
Mexico	2750-0942F(MX)	00038.102	6/16/00	
United States	2750-0948P	80132.017	6/16/00	09/595,329
Mexico	2750-0943F(MX)	00039.102	6/16/00	
United States	2750-0947P	80132.016	6/16/00	09/595,335
United States	2750-0945P	80132.014	6/16/00	09/595,333
United States	2750-0944P	80132.013	6/16/00	09/595,330
United States	2750-0943P	00039.002	6/16/00	09/596,577
United States	2750-0954P	80132.023	6/16/00	09/594,599
United States	2750-0950P	80132.019	6/16/00	09/594,598
Canada	2750-0941F(CA)	00037.100	6/16/00	
United States	2750-0953P	80132.022	6/16/00	09/595,298
United States	2750-0949P	80132.018	6/16/00	09/595,332
United States	2750-0941P	00037.002	6/16/00	09/595,334
United States	2750-0951P	80132.020	6/16/00	09/594,595
United States	2750-0946P	80132.015	6/16/00	09/595,328
Canada	2750-0942F(CA)	00038.100	6/16/00	
United States	2750-0942P	00038.002	6/16/00	09/595,326
United States	2750-0952P	80132.021	6/16/00	09/594,597
United States	2750-0958P	91075.001	6/19/00	60/212,623
United States	2750-0957P	80275.001	6/19/00	60/212,649
United States	2750-0956P	00239.001	6/19/00	60/212,414

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0960P	80276.001	6/20/00	60/212,713
United States	2750-0959P	00240.001	6/20/00	60/212,677
United States	2750-0961P	91076.001	6/20/00	60/212,727
United States	2750-0971P	00042.003	6/21/00	09/602,660
Canada	2750-0971F(CA)	00042.100	6/21/00	02/309,874
Europe	2750-0971F(EP)	00042.101	6/21/00	00/305,249
Mexico	2750-0971F(MX)	00042.102	6/21/00	00/006,142
United States	2750-0967P	91079.001	6/22/00	60/213,270
United States	2750-0966P	80278.001	6/22/00	60/213,220
United States	2750-0965P	00246.001	6/22/00	60/213,221
Mexico	2750-0972F(MX)	00043.102	6/22/00	00/006,625
Europe	2750-0972F(EP)	00043.101	6/22/00	00/305,270
Canada	2750-0972F(CA)	00043.100	6/22/00	02/309,793
United States	2750-0972P	00043.002	6/22/00	09/602,152
United States	2750-0963P	80277.001	6/22/00	60/213,249
United States	2750-0962P	00242.001	6/22/00	60/213,271
United States	2750-0964P	91077.001	6/22/00	60/213,195
Mexico	2750-0974F(MX)	00042.102	6/23/00	
Mexico	2750-0975F(MX)	00045.102	6/23/00	
Europe	2750-0973F(EP)	00044.101	6/23/00	00/305,305
Canada	2750-0975F(CA)	00045.100	6/23/00	
Europe	2750-0974F(EP)	00042.101	6/23/00	
Mexico	2750-0973F(MX)	00044.102	6/23/00	00/006,267
United States	2750-0973P	00044.002	6/23/00	09/602,205
Canada	2750-0973F(CA)	00044.100	6/23/00	02/309,889
Europe	2750-0975F(EP)	00045.101	6/23/00	
United States	2750-0975P	00045.002	6/23/00	09/602,016
Canada	2750-0974F(CA)	00042.100	6/23/00	
United States	2750-1035P	00248.001	6/27/00	60/214,534
United States	2750-0968P	00247.001	6/27/00	
United States	2750-0969P	80279.001	6/27/00	60/214,762
United States	2750-0970P	91080.001	6/27/00	60/214,524
United States	2750-1036P	80280.001	6/27/00	60/214,535
Canada	2750-0976F(CA)	00046.100	6/28/00	
Europe	2750-0976F(EP)	00046.101	6/28/00	
Mexico	2750-0976F(MX)	00046.102	6/28/00	
United States	2750-1038P	00249.001	6/28/00	60/214,800
United States	2750-1039P	80281.001	6/28/00	60/214,799
United States	2750-0976P	00046.002	6/28/00	09/605,843
Canada	2750-0977F(CA)	00048.100	6/29/00	
Europe	2750-0977F(EP)	00048.101	6/29/00	
Mexico	2750-0977F(MX)	00048.102	6/29/00	
United States	2750-0977P	00048.002	6/29/00	09/606,181
Europe	2750-0981F(EP)	00052.101	6/30/00	
United States	2750-0980P	00051.002	6/30/00	09/610,157
Canada	2750-0979F(CA)	00050.100	6/30/00	
Europe	2750-0979F(EP)	00050.101	6/30/00	
Mexico	2750-0979F(MX)	00050.102	6/30/00	
United States	2750-1041P	80282.001	6/30/00	60/215,127
United States	2750-1040P	00250.001	6/30/00	60/215,775

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-0979P	00050.002	6/30/00	09/607,081
United States	2750-0981P	00052.002	6/30/00	09/609,198
Mexico	2750-0980F(MX)	00051.102	6/30/00	
Canada	2750-0980F(CA)	00051.100	6/30/00	
Mexico	2750-0978F(MX)	00049.102	6/30/00	
Europe	2750-0978F(EP)	00049.101	6/30/00	
Canada	2750-0978F(CA)	00049.100	6/30/00	
Canada	2750-0981F(CA)	00052.100	6/30/00	
Europe	2750-0980F(EP)	00051.101	6/30/00	
United States	2750-0978P	00049.002	6/30/00	09/608,960
Mexico	2750-0981F(MX)	00052.102	6/30/00	
United States	2750-1043P	80283.001	7/5/00	60/216,362
United States	2750-1042P	00252.001	7/5/00	
Mexico	2750-0982F(MX)	00053.102	7/6/00	
United States	2750-0982P	00053.002	7/6/00	09/611,409
Europe	2750-0982F(EP)	00053.101	7/6/00	
Canada	2750-0982F(CA)	00053.100	7/6/00	
Mexico	2750-0983F(MX)	00054.102	7/7/00	
United States	2750-0984P	00058.002	7/7/00	09/613,547
Europe	2750-0984F(EP)	00058.101	7/7/00	
Mexico	2750-0984F(MX)	00058.102	7/7/00	
United States	2750-0983P	00054.002	7/7/00	09/612,645
Canada	2750-0984F(CA)	00058.100	7/7/00	
Europe	2750-0983F(EP)	00054.101	7/7/00	
Canada	2750-0983F(CA)	00054.100	7/7/00	
United States	2750-1044P	91081.001	7/11/00	60/217,385
United States	2750-1045P	00253.001	7/11/00	60/217,476
United States	2750-1046P	80284.001	7/11/00	60/217,384
United States	2750-0985P	00059.002	7/12/00	09/615,007
Canada	2750-0985F(CA)	00059.100	7/12/00	
Europe	2750-0985F(EP)	00059.101	7/12/00	
United States	2750-1050P	80286.002	7/12/00	
Mexico	2750-0985F(MX)	00059.102	7/12/00	
United States	2750-1052P	80287.002	7/12/00	
United States	2750-0986P	00060.002	7/13/00	09/615,748
Canada	2750-0986F(CA)	00060.100	7/13/00	
United States	2750-1054P	80288.002	7/13/00	
Mexico	2750-0986F(MX)	00060.102	7/13/00	
Europe	2750-0986F(EP)	00060.101	7/13/00	
Canada	2750-0988F(CA)	00062.100	7/14/00	
United States	2750-1060P	80134.017	7/14/00	09/614,388
United States	2750-0987P	00061.002	7/14/00	09/617,525
Canada	2750-0987F(CA)	00061.100	7/14/00	
Europe	2750-0987F(EP)	00061.101	7/14/00	
United States	2750-1061P	80134.018	7/14/00	09/614,450
United States	2750-0988P	00062.002	7/14/00	09/617,203
Europe	2750-0988F(EP)	00062.101	7/14/00	
Mexico	2750-0988F(MX)	00062.102	7/14/00	
United States	2750-1048P	80285.002	7/14/00	
Mexico	2750-0987F(MX)	00061.102	7/14/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-1055P	91082.001	7/18/00	
United States	2750-1057P	80291.001	7/18/00	
United States	2750-1056P	00254.001	7/18/00	
Mexico	2750-0989F(MX)	00064.102	7/19/00	
Europe	2750-0989F(EP)	00064.101	7/19/00	
United States	2750-0989P	00064.002	7/19/00	09/620,421
United States	2750-1062P	80134.020	7/19/00	09/617,683
United States	2750-1063P	80134.022	7/19/00	09/617,682
United States	2750-1064P	80134.024	7/19/00	09/617,681
Canada	2750-0989F(CA)	00064.100	7/19/00	
Canada	2750-0990F(CA)	00065.100	7/20/00	
United States	2750-0990P	00065.002	7/20/00	09/621,323
Europe	2750-0990F(EP)	00065.101	7/20/00	
Mexico	2750-0990F(MX)	00065.102	7/20/00	
United States	2750-1065P	80134.026	7/20/00	09/620,998
United States	2750-1066P	80135.004	7/20/00	09/620,978
United States	2750-1073P	80134.025	7/21/00	09/620,314
United States	2750-1072P	80135.003	7/21/00	09/621,900
United States	2750-0992P	00067.002	7/21/00	09/621,660
Canada	2750-0992F(CA)	00067.100	7/21/00	
Europe	2750-0992F(EP)	00067.101	7/21/00	
United States	2750-1069P	80134.023	7/21/00	09/620,313
United States	2750-1071P	80134.021	7/21/00	09/620,390
United States	2750-1070P	80134.019	7/21/00	09/620,111
United States	2750-1068P	80134.016	7/21/00	09/620,393
Mexico	2750-0993F(MX)	00069.102	7/21/00	
Europe	2750-0993F(EP)	00069.101	7/21/00	
Canada	2750-0993F(CA)	00069.100	7/21/00	
United States	2750-0991P	00066.002	7/21/00	09/621,630
Canada	2750-0991F(CA)	00066.100	7/21/00	
United States	2750-0993P	00069.002	7/21/00	09/621,902
Europe	2750-0991F(EP)	00066.101	7/21/00	
Mexico	2750-0991F(MX)	00066.102	7/21/00	
Mexico	2750-0992F(MX)	00067.102	7/21/00	
United States	2750-1067P	80134.015	7/21/00	09/620,394
United States	2750-1059P	00255.001	7/25/00	
United States	2750-1079P	80292.001	7/25/00	
United States	2750-1081P	80293.001	7/25/00	
United States	2750-1058P	91083.001	7/25/00	
United States	2750-1080P	00256.001	7/25/00	
Canada	2750-0994F(CA)	00070.100	7/26/00	
Europe	2750-0994F(EP)	00070.101	7/26/00	
Mexico	2750-0994F(MX)	00070.102	7/26/00	
United States	2750-0994P	00070.002	7/26/00	
United States	2750-1075P	80136.005	7/27/00	09/628,987
United States	2750-0995P	00071.002	7/27/00	09/628,986
Canada	2750-0995F(CA)	00071.100	7/27/00	
United States	2750-1074P	80136.004	7/27/00	09/628,984
Europe	2750-0995F(EP)	00071.101	7/27/00	
Mexico	2750-0995F(MX)	00071.102	7/27/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
Europe	2750-0996F(EP)	00072.101	7/28/00	
Canada	2750-0996F(CA)	00072.100	7/28/00	
Mexico	2750-0996F(MX)	00072.102	7/28/00	
United States	2750-0996P	00072.002	7/28/00	09/628,552
United States	2750-1049P	80286.001	8/1/00	
United States	2750-1256P	80299.002	8/1/00	
United States	2750-1077P	80137.004	8/2/00	09/630,442
Mexico	2750-0997F(MX)	00073.102	8/2/00	
Europe	2750-0997F(EP)	00073.101	8/2/00	
Canada	2750-0997F(CA)	00073.100	8/2/00	
United States	2750-0997P	00073.002	8/2/00	09/632,340
United States	2750-1076P	80137.003	8/2/00	09/628,985
United States	2750-1257P	80299.003	8/3/00	
United States	2750-1051P	80287.001	8/3/00	
United States	2750-0998P	00074.002	8/3/00	09/632,349
United States	2750-1258P	80299.004	8/3/00	
United States	2750-1053P	80288.002	8/3/00	60/223,100
United States	2750-1255P	80299.001	8/3/00	
Mexico	2750-0998F(MX)	00074.102	8/3/00	
Europe	2750-0998F(EP)	00074.101	8/3/00	
Canada	2750-0998F(CA)	00074.100	8/3/00	
United States	2750-1078P	80138.003	8/4/00	09/635,640
Europe	2750-1000F(EP)	00077.101	8/4/00	
Canada	2750-1000F(CA)	00077.100	8/4/00	
United States	2750-1000P	00077.002	8/4/00	09/633,191
Mexico	2750-1001F(MX)	00079.102	8/4/00	
Mexico	2750-1000F(MX)	00077.102	8/4/00	
Europe	2750-1001F(EP)	00079.101	8/4/00	
United States	2750-1001P	00079.002	8/4/00	09/633,239
Mexico	2750-0999F(MX)	00076.102	8/4/00	
United States	2750-1092P	80138.004	8/4/00	09/635,643
Canada	2750-1001F(CA)	00079.100	8/4/00	
United States	2750-0999P	00076.002	8/4/00	09/633,051
Canada	2750-0999F(CA)	00076.100	8/4/00	
Europe	2750-0999F(EP)	00076.101	8/4/00	
United States	2750-1047P	80285.001	8/7/00	
United States	2750-1094P	80139.004	8/9/00	09/635,642
United States	2750-1114P	92001.004	8/9/00	60/224,390
Europe	2750-1002F(EP)	00080.101	8/9/00	
United States	2750-1115P	92002.004	8/9/00	
Mexico	2750-1002F(MX)	00080.102	8/9/00	
Canada	2750-1002F(CA)	00080.100	8/9/00	
United States	2750-1002P	00080.002	8/9/00	09/635,277
United States	2750-1093P	80139.003	8/10/00	09/635,641
Canada	2750-1005F(CA)	00083.100	8/11/00	
United States	2750-1101P	80141.005	8/11/00	
United States	2750-1005P	00083.002	8/11/00	09/637,563
United States	2750-1096P	80142.004	8/11/00	09/637,780
United States	2750-1103P	80141.007	8/11/00	
Europe	2750-1005F(EP)	00083.101	8/11/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
Canada	2750-1003F(CA)	00081.100	8/11/00	
United States	2750-1003P	00081.002	8/11/00	09/637,837
Mexico	2750-1005F(MX)	00083.102	8/11/00	
United States	2750-1004P	00082.002	8/11/00	09/636,555
Canada	2750-1004F(CA)	00082.100	8/11/00	
Europe	2750-1004F(EP)	00082.101	8/11/00	
Mexico	2750-1004F(MX)	00082.102	8/11/00	
United States	2750-1104P	80141.008	8/11/00	
Europe	2750-1003F(EP)	00081.101	8/11/00	
Mexico	2750-1003F(MX)	00081.102	8/11/00	
United States	2750-1102P	80141.006	8/11/00	
United States	2750-1083P	91084.001	8/14/00	
United States	2750-1082P	80298.001	8/14/00	
United States	2750-1084P	80300.001	8/15/00	
United States	2750-1085P	91085.001	8/15/00	
United States	2750-1106P	80294.002	8/16/00	
United States	2750-1110P	80296.002	8/16/00	
United States	2750-1108P	80295.002	8/16/00	
United States	2750-1112P	80297.002	8/16/00	
United States	2750-1095P	80142.003	8/16/00	09/643,672
United States	2750-1006P	00084.002	8/17/00	09/641,198
Europe	2750-1006F(EP)	00084.101	8/17/00	
Mexico	2750-1006F(MX)	00084.102	8/17/00	
Canada	2750-1006F(CA)	00084.100	8/17/00	
Mexico	2750-1007F(MX)	00085.102	8/18/00	
United States	2750-1111P	80297.001	8/18/00	60/226,324
Europe	2750-1007F(EP)	00085.101	8/18/00	
United States	2750-1109P	80296.001	8/18/00	
United States	2750-1107P	80295.001	8/18/00	
Canada	2750-1007F(CA)	00085.100	8/18/00	
United States	2750-1007P	00085.002	8/18/00	
Canada	2750-1009F(CA)	00087.100	8/18/00	
Mexico	2750-1008F(MX)	00086.102	8/18/00	
United States	2750-1105P	80294.001	8/18/00	
mexico	2750-1009F(MX)	00087.102	8/18/00	
Europe	2750-1009F(EP)	00087.101	8/18/00	
United States	2750-1098P	80143.004	8/18/00	09/643,671
United States	2750-1008P	00086.002	8/18/00	
Canada	2750-1008F(CA)	00086.100	8/18/00	
Europe	2750-1008F(EP)	00086.101	8/18/00	
United States	2750-1009P	00087.002	8/18/00	09/640,695
United States	2750-1086P	80301.001	8/21/00	
United States	2750-1087P	91086.001	8/21/00	
United States	2750-1117P	80289.002	8/23/00	
United States	2750-1118P	80289.003	8/23/00	
United States	2750-1122P	80289.007	8/23/00	
United States	2750-1132P	80289.017	8/23/00	
United States	2750-1124P	80289.009	8/23/00	
United States	2750-1131P	80289.016	8/23/00	
United States	2750-1116P	80289.001	8/23/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-1123P	80289.008	8/23/00	
Europe	2750-1010F(EP)	00088.101	8/23/00	
United States	2750-1010P	00088.002	8/23/00	09/643,854
Mexico	2750-1010F(MX)	00088.102	8/23/00	
United States	2750-1162P	80307.001	8/23/00	
United States	2750-1163P	91087.001	8/23/00	
United States	2750-1097P	80143.003	8/23/00	09/649,866
Canada	2750-1010F(CA)	00088.100	8/23/00	
Mexico	2750-1013F(MX)	00091.102	8/25/00	
United States	2750-1099P	80144.003	8/25/00	09/649,868
United States	2750-1133P	80289.018	8/25/00	
United States	2750-1140P	80289.025	8/25/00	
United States	2750-1145P	80289.030	8/25/00	60/227,728
United States	2750-1150P	80289.035	8/25/00	
United States	2750-1151P	80289.036	8/25/00	
United States	2750-1149P	80289.034	8/25/00	
United States	2750-1146P	80289.031	8/25/00	
United States	2750-1147P	80289.032	8/25/00	
United States	2750-1144P	80289.029	8/25/00	
United States	2750-1143P	80289.028	8/25/00	
United States	2750-1137P	80289.022	8/25/00	
United States	2750-1136P	80289.021	8/25/00	
United States	2750-1135P	80289.020	8/25/00	
United States	2750-1134P	80289.019	8/25/00	
United States	2750-1120P	80289.005	8/25/00	
United States	2750-1013P	00091.002	8/25/00	
Canada	2750-1013F(CA)	00091.100	8/25/00	
United States	2750-1152P	80289.037	8/25/00	
Mexico	2750-1012F(MX)	00090.102	8/25/00	
United States	2750-1126P	80289.011	8/25/00	
Europe	2750-1013F(EP)	00091.101	8/25/00	
United States	2750-1011P	00089.002	8/25/00	
Canada	2750-1011F(CA)	00089.100	8/25/00	
Europe	2750-1011F(EP)	00089.101	8/25/00	
Mexico	2750-1011F(MX)	00089.102	8/25/00	
United States	2750-1138P	80289.023	8/25/00	60/227,729
United States	2750-1119P	80289.004	8/25/00	60/228,029
United States	2750-1100P	80144.004	8/25/00	09/649,867
United States	2750-1148P	80289.033	8/25/00	
United States	2750-1142P	80289.027	8/25/00	60/227,733
United States	2750-1128P	80289.013	8/25/00	60/227,726
United States	2750-1139P	80289.024	8/25/00	60/228,167
United States	2750-1121P	80289.006	8/25/00	
Europe	2750-1012F(EP)	00090.101	8/25/00	
Canada	2750-1012F(CA)	00090.100	8/25/00	
United States	2750-1012P	00090.002	8/25/00	
United States	2750-1141P	80289.026	8/25/00	
United States	2750-1127P	80289.012	8/25/00	
United States	2750-1225P	80308.001	8/30/00	
Mexico	2750-1014F(MX)	00092.102	8/30/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
Europe	2750-1014F(EP)	00092.101	8/30/00	
Canada	2750-1014F(CA)	00092.100	8/30/00	
United States	2750-1226P	91088.001	8/30/00	
United States	2750-1014P	00092.002	8/30/00	09/651,370
United States	2750-1090P	80303.001	8/31/00	
United States	2750-1089P	80302.002	8/31/00	
United States	2750-1088P	80302.001	8/31/00	
Mexico	2750-1015F(MX)	00093.102	8/31/00	
United States	2750-1015P	00093.002	8/31/00	09/653,466
United States	2750-1091P	80303.002	8/31/00	
Canada	2750-1015F(CA)	00093.100	8/31/00	
Europe	2750-1015F(EP)	00093.101	8/31/00	
United States	2750-1113P	80304.001	8/31/00	
Mexico	2750-1016F(MX)	00094.102	9/1/00	
United States	2750-1016P	00094.002	9/1/00	09/654,547
Europe	2750-1016F(EP)	00094.101	9/1/00	
Canada	2750-1016F(CA)	00094.100	9/1/00	
United States	2750-1228P	91089.001	9/6/00	
United States	2750-1227P	80309.001	9/6/00	
United States	2750-1160P	80306.001	9/6/00	
United States	2750-1159P	80305.002	9/6/00	
United States	2750-1158P	80305.001	9/6/00	
United States	2750-1157P	80304.002	9/6/00	
United States	2750-1161P	80306.002	9/6/00	
Mexico	2750-1017F(MX)	00095.102	9/7/00	
Canada	2750-1017F(CA)	00095.100	9/7/00	
United States	2750-1017P	00095.002	9/7/00	
Europe	2750-1017F(EP)	00095.101	9/7/00	
Mexico	2750-1018F(MX)	00096.102	9/8/00	
Europe	2750-1018F(EP)	00096.101	9/8/00	
Canada	2750-1018F(CA)	00096.100	9/8/00	
United States	2750-1018P	00096.002	9/8/00	09/657,569
United States	2750-1230P	91090.001	9/13/00	
Europe	2750-1019F(EP)	00098.101	9/13/00	
Canada	2750-1019F(CA)	00098.100	9/13/00	
United States	2750-1019P	00098.002	9/13/00	09/660,883
United States	2750-1229P	80310.001	9/13/00	
Mexico	2750-1019F(MX)	00098.102	9/13/00	
United States	2750-1231P	80311.001	9/15/00	
United States	2750-1021P	00101.002	9/15/00	
Canada	2750-1021F(CA)	00101.100	9/15/00	
Europe	2750-1021F(EP)	00101.101	9/15/00	
Mexico	2750-1021F(MX)	00101.102	9/15/00	
Canada	2750-1020F(CA)	00099.100	9/15/00	
United States	2750-1232P	91091.001	9/15/00	
Europe	2750-1020F(EP)	00099.101	9/15/00	
Mexico	2750-1020F(MX)	00099.102	9/15/00	
United States	2750-1020P	00099.002	9/15/00	
United States	2750-1233P	80312.001	9/18/00	
United States	2750-1234P	91092.001	9/18/00	

Country	Attorney Docket	Client Docket	Filed	Application No.
United States	2750-1253P	80313.001	9/20/00	
Mexico	2750-1022F(MX)	00102.102	9/20/00	
Europe	2750-1022F(EP)	00102.101	9/20/00	
United States	2750-1022P	00102.002	9/20/00	
Canada	2750-1022F(CA)	00102.100	9/20/00	
United States	2750-1254P	91093.001	9/20/00	
United States	2750-1259P	80314.001	9/21/00	
United States	2750-1260P	91094.001	9/21/00	
United States	2750-1261P	91095.001	9/21/00	
Mexico	2750-1023F(MX)	00103.102	9/22/00	
Mexico	2750-1025F(MX)	00105.102	9/22/00	
United States	2750-1025P	00105.002	9/22/00	
Europe	2750-1023F(EP)	00103.101	9/22/00	
Canada	2750-1025F(CA)	00105.100	9/22/00	
Canada	2750-1023F(CA)	00103.100	9/22/00	
United States	2750-1023P	00103.002	9/22/00	
Mexico	2750-1024F(MX)	00104.102	9/22/00	
Europe	2750-1024F(EP)	00104.101	9/22/00	
Canada	2750-1024F(CA)	00104.100	9/22/00	
United States	2750-1024P	00104.002	9/22/00	
Europe	2750-1025F(EP)	00105.101	9/22/00	
United States	2750-1263P	91096.001	9/25/00	
United States	2750-1264P	80316.001	9/25/00	
United States	2750-1265P	91097.001	9/25/00	
United States	2750-1262P	80315.001	9/25/00	
United States	2750-1266P	80317.001	9/26/00	
United States	2750-1267P	91098.001	9/27/00	
United States	2750-1026P	00106.002	9/28/00	
United States	2750-1268P	91078.001	9/28/00	
United States	2750-1027P	00107.002	9/29/00	
United States	2750-1269P	91099.001	9/29/00	
United States	2750-1271P	91100.001	10/2/00	
United States	2750-1270P	80318.001	10/2/00	
United States	2750-1028P	00108.002	10/4/00	
United States	2750-1273P	91101.001	10/4/00	
United States	2750-1272P	80319.001	10/4/00	
United States	2750-1029P	00109.002	10/5/00	
United States	2750-1274P	80320.001	10/5/00	
United States	2750-1032P	00112.002	10/6/00	
United States	2750-1031P	00111.002	10/6/00	
United States	2750-1030P	00110.002	10/6/00	

All applications listed in the table above are expressly incorporated herein by reference.

The SDFs of the invention can also be used as probes to search for genes that are related to the SDF within a species. Such related genes are typically considered to be members of a gene family. In such a case, the sequence similarity will often be concentrated

into one or a few fragments of the sequence. The fragments of similar sequence that define the gene family typically encode a fragment of a protein or RNA that has an enzymatic or structural function. The percentage of identity in the amino acid sequence of the domain that defines the gene family is preferably at least 70%, more preferably 80 to 95%, most preferably 85 to 99%. To search for members of a gene family within a species, a low stringency hybridization is usually performed, but this will depend upon the size, distribution and degree of sequence divergence of domains that define the gene family. SDFs encompassing regulatory regions can be used to identify coordinately expressed genes by using the regulatory region sequence of the SDF as a probe.

In the instances where the SDFs are identified as being expressed from genes that confer a particular phenotype, then the SDFs can also be used as probes to assay plants of different species for those phenotypes.

I.C. Methods to Inhibit Gene Expression

The nucleic acid molecules of the present invention can be used to inhibit gene transcription and/or translation. Example of such methods include, without limitation:

Antisense Constructs;

Ribozyme Constructs;

Chimeraplast Constructs;

Co-Suppression;

Transcriptional Silencing; and

Other Methods of Gene Expression.

C.1 <u>Antisense</u>

In some instances it is desirable to suppress expression of an endogenous or exogenous gene. A well-known instance is the FLAVOR-SAVOR™ tomato, in which the gene encoding ACC synthase is inactivated by an antisense approach, thus delaying softening of the fruit after ripening. See for example, U.S. Patent No. 5,859,330; U.S. Patent No. 5,723,766; Oeller, et al, *Science,* 254:437-439(1991); and Hamilton et al, *Nature*, 346:284-287 (1990). Also, timing of flowering can be controlled by suppression of the *FLOWERING LOCUS C* (*FLC*); high levels of this transcript are associated with late flowering, while absence of *FLC* is associated with early flowering (S.D. Michaels et al., *Plant Cell* <u>11</u>:949 (1999). Also, the transition of apical meristem from production of leaves with associated

shoots to flowering is regulated by *TERMINAL FLOWER1, APETALA1* and *LEAFY*. Thus, when it is desired to induce a transition from shoot production to flowering, it is desirable to suppress *TFL1* expression (S.J. Liljegren, *Plant Cell* 11:1007 (1999)). As another instance, arrested ovule development and female sterility result from suppression of the ethylene forming enzyme but can be reversed by application of ethylene (D. De Martinis et al., *Plant Cell* 11:1061 (1999)). The ability to manipulate female fertility of plants is useful in increasing fruit production and creating hybrids.

In the case of polynucleotides used to inhibit expression of an endogenous gene, the introduced sequence need not be perfectly identical to a sequence of the target endogenous gene. The introduced polynucleotide sequence will typically be at least substantially identical to the target endogenous sequence.

Some polynucleotide SDFs in Tables 1 and 2 represent sequences that are expressed in corn,wheat, rice, soybean *Arabidopsis* and/or other plants. Thus the invention includes using these sequences to generate antisense constructs to inhibit translation and/or degradation of transcripts of said SDFs, typically in a plant cell.

To accomplish this, a polynucleotide segment from the desired gene that can hybridize to the mRNA expressed from the desired gene (the antisense segment") is operably linked to a promoter such that the antisense strand of RNA will be transcribed when the construct is present in a host cell. A regulated promoter can be used in the construct to control transcription of the antisense segment so that transcription occurs only under desired circumstances.

The antisense segment to be introduced generally will be substantially identical to at least a fragment of the endogenous gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. Further, the antisense product may hybridize to the untranslated region instead of or in addition to the coding sequence of the gene. The vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced antisense segment sequence also need not be full length relative to either the primary transcription product or the fully processed mRNA. Generally, a higher percentage of sequence identity can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and the full length of the transcript canbe used, though a sequence of at least about 100 nucleotides is preferred, a

sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

C.2. Ribozymes

It is also contemplated that gene constructs representing ribozymes and based on the SDFs in Tables 1 and 2 are an object of the invention. Ribozymes can also be used to inhibit expression of genes by suppressing the translation of the mRNA into a polypeptide. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs, which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature,* 334:585 (1988).

Like the antisense constructs above, the ribozyme sequence fragment necessary for pairing need not be identical to the target nucleotides to be cleaved, nor identical to the sequences in Tables 1 and 2. Ribozymes may be constructed by combining the ribozyme sequence and some fragment of the target gene which would allow recognition of the target gene mRNA by the resulting ribozyme molecule. Generally, the sequence in the ribozyme capable of binding to the target sequence exhibits a percentage of sequence identity with at least 80%, preferably with at least 85%, more preferably with at least 90% and most preferably with at least 95%, even more preferably, with at least 96%, 97%, 98% or 99% sequence identity to some fragment of a sequence in Tables 1 and 2 or the complement thereof. The ribozyme can be equally effective in inhibiting mRNA translation by cleaving either in the untranslated or coding regions. Generally, a higher percentage of sequence identity can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not

have the same intron or exon pattern, and homology of non-coding segments may be equally effective.

C.3. Chimeraplasts

The SDFs of the invention, such as those described by Tables 1 and 2, can also be used to construct chimeraplasts that can be introduced into a cell to produce at least one specific nucleotide change in a sequence corresponding to the SDF of the invention. A chimeraplast is an oligonucleotide comprising DNA and/or RNA that specifically hybridizes to a target region in a manner which creates a mismatched base-pair. This mismatched base-pair signals the cell's repair enzyme machinery which acts on the mismatched region resulting in the replacement, insertion or deletion of designated nucleotide(s). The altered sequence is then expressed by the cell's normal cellular mechanisms. Chimeraplasts can be designed to repair mutant genes, modify genes, introduce site-specific mutations, and/or act to interrupt or alter normal gene function (US Pat. Nos. 6,010,907 and 6,004,804; and PCT Pub. No. WO99/58723 and WO99/07865).

C.4. <u>Sense Suppression</u>

The SDFs of Tables 1 and 2 of the present invention are also useful to modulate gene expression by sense suppression. Sense suppression represents another method of gene suppression by introducing at least one exogenous copy or fragment of the endogenous sequence to be suppressed.

Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter into the chromosome of a plant or by a self-replicating virus has been shown to be an effective means by which to induce degradation of mRNAs of target genes. For an example of the use of this method to modulate expression of endogenous genes *see*, Napoli et al., *The Plant Cell* 2:279 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184. Inhibition of expression may require some transcription of the introduced sequence.

For sense suppression, the introduced sequence generally will be substantially identical to the endogenous sequence intended to be inactivated. The minimal percentage of sequence identity will typically be greater than about 65%, but a higher percentage of sequence identity might exert a more effective reduction in the level of normal gene products. Sequence identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect would likely apply to any other proteins

within a similar family of genes exhibiting homology or substantial homology to the suppressing sequence.

C.5. Transcriptional Silencing

The nucleic acid sequences of the invention, including the SDFs of Tables 1 and 2, and fragments thereof, contain sequences that can be inserted into the genome of an organism resulting in transcriptional silencing. Such regulatory sequences need not be operatively linked to coding sequences to modulate transcription of a gene. Specifically, a promoter sequence without any other element of a gene can be introduced into a genome to transcriptionally silence an endogenous gene (see, for example, Vaucheret, H et al. (1998) The Plant Journal 16: 651-659). As another example, triple helices can be formed using oligonucleotides based on sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto. The oligonucleotide can be delivered to the host cell and can bind to the promoter in the genome to form a triple helix and prevent transcription. An oligonucleotide of interest is one that can bind to the promoter and block binding of a transcription factor to the promoter. In such a case, the oligonucleotide can be complementary to the sequences of the promoter that interact with transcription binding factors.

C.6. Other Methods to Inhibit Gene Expression

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Low frequency homologous recombination can be used to target a polynucleotide insert to a gene by flanking the polynucleotide insert with sequences that are substantially similar to the gene to be disrupted. Sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto can be used for homologous recombination.

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred to identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or R_1 plants having a desired phenotype.

I.D. Methods of Functional Analysis

The constructs described in the methods under I.C. above can be used to determine the function of the polypeptide encoded by the gene that is targeted by the constructs.

Down-regulating the transcription and translation of the targeted gene in the host cell or organisms, such as a plant, may produce phenotypic changes as compared to a wild-type cell or organism. In addition, *in vitro* assays can be used to determine if any biological activity, such as calcium flux, DNA transcription, nucleotide incorporation, etc., are being modulated by the down-regulation of the targeted gene.

Coordinated regulation of sets of genes, e.g., those contributing to a desired polygenic trait, is sometimes necessary to obtain a desired phenotype. SDFs of the invention representing transcription activation and DNA binding domains can be assembled into hybrid transcriptional activators. These hybrid transcriptional activators can be used with their corresponding DNA elements (i.e., those bound by the DNA-binding SDFs) to effect coordinated expression of desired genes (J.J. Schwarz et al., *Mol. Cell. Biol.* 12:266 (1992), A. Martinez et al., *Mol. Gen. Genet.* 261:546 (1999)).

The SDFs of the invention can also be used in the two-hybrid genetic systems to identify networks of protein-protein interactions (L. McAlister-Henn et al., *Methods* 19:330 (1999), J.C. Hu et al., *Methods* 20:80 (2000), M. Golovkin et al., *J. Biol. Chem.* 274:36428 (1999), K. Ichimura et al., *Biochem. Biophys. Res. Comm.* 253:532 (1998)). The SDFs of the invention can also be used in various expression display methods to identify important protein-DNA interactions (e.g. B. Luo et al., *J. Mol. Biol.* 266:479 (1997)).

I.E. Promoters

The SDFs of the invention are also useful as structural or regulatory sequences in a construct for modulating the expression of the corresponding gene in a plant or other organism, *e.g.* a symbiotic bacterium. For example, promoter sequences associated to SDFs of Tables 1 and 2 of the present invention can be useful in directing expression of coding sequences either as constitutive promoters or to direct expression in particular cell types, tissues, or organs or in response to environmental stimuli.

With respect to the SDFs of the present invention a promoter is likely to be a relatively small portion of a genomic DNA (gDNA) sequence located in the first 2000 nucleotides

upstream from an initial exon identified in a gDNA sequence or initial ATG" or methionine codon or translational start site in a corresponding cDNA sequence. Such promoters are more likely to be found in the first 1000 nucleotides upstream of an initial ATG or methionine codon or translational start site of a cDNA sequence corresponding to a gDNA sequence. In particular, the promoter is usually located upstream of the transcription start site. The fragments of a particular gDNA sequence that function as elements of a promoter in a plant cell will preferably be found to hybridize to gDNA sequences presented and described in Tables 1 and 2 at medium or high stringency, relevant to the length of the probe and its base composition.

Promoters are generally modular in nature. Promoters can consist of a basal promoter that functions as a site for assembly of a transcription complex comprising an RNA polymerase, for example RNA polymerase II. A typical transcription complex will include additional factors such as $TF_{II}B$, $TF_{II}D$, and $TF_{II}E$. Of these, $TF_{II}D$ appears to be the only one to bind DNA directly. The promoter might also contain one or more enhancers and/or suppressors that function as binding sites for additional transcription factors that have the function of modulating the level of transcription with respect to tissue specificity and of transcriptional responses to particular environmental or nutritional factors, and the like.

Short DNA sequences representing binding sites for proteins can be separated from each other by intervening sequences of varying length. For example, within a particular functional module, protein binding sites may be constituted by regions of 5 to 60, preferably 10 to 30, more preferably 10 to 20 nucleotides. Within such binding sites, there are typically 2 to 6 nucleotides that specifically contact amino acids of the nucleic acid binding protein. The protein binding sites are usually separated from each other by 10 to several hundred nucleotides, typically by 15 to 150 nucleotides, often by 20 to 50 nucleotides. DNA binding sites in promoter elements often display dyad symmetry in their sequence. Often elements binding several different proteins, and/or a plurality of sites that bind the same protein, will be combined in a region of 50 to 1,000 basepairs.

Elements that have transcription regulatory function can be isolated from their corresponding endogenous gene, or the desired sequence can be synthesized, and recombined in constructs to direct expression of a coding region of a gene in a desired tissue-specific, temporal-specific or other desired manner of inducibility or suppression. When hybridizations are performed to identify or isolate elements of a promoter by hybridization to the long sequences presented in Tables 1 and 2, conditions are adjusted to account for the above-described nature of promoters. For example short probes, constituting the element sought, are preferably used under low temperature and/or high salt conditions. When long probes, which might include several

promoter elements are used, low to medium stringency conditions are preferred when hybridizing to promoters across species.

If a nucleotide sequence of an SDF, or part of the SDF, functions as a promoter or fragment of a promoter, then nucleotide substitutions, insertions or deletions that do not substantially affect the binding of relevant DNA binding proteins would be considered equivalent to the exemplified nucleotide sequence. It is envisioned that there are instances where it is desirable to decrease the binding of relevant DNA binding proteins to silence or down-regulate a promoter, or conversely to increase the binding of relevant DNA binding proteins to enhance or up-regulate a promoter and vice versa. In such instances, polynucleotides representing changes to the nucleotide sequence of the DNA-protein contact region by insertion of additional nucleotides, changes to identity of relevant nucleotides, including use of chemically-modified bases, or deletion of one or more nucleotides are considered encompassed by the present invention. In addition, fragments of the promoter sequences described by Tables 1 and 2 and variants thereof can be fused with other promoters or fragments to facilitate transcription and/or transcription in specific type of cells or under specific conditions.

Promoter function can be assayed by methods known in the art, preferably by measuring activity of a reporter gene operatively linked to the sequence being tested for promoter function. Examples of reporter genes include those encoding luciferase, green fluorescent protein, GUS, neo, cat and bar.

I.F. UTRs and Junctions

Polynucleotides comprising untranslated (UTR) sequences and intron/exon junctions are also within the scope of the invention. UTR sequences include introns and 5' or 3' untranslated regions (5' UTRs or 3' UTRs). Fragments of the sequences shown in Tables 1 and 2 can comprise UTRs and intron/exon junctions.

These fragments of SDFs, especially UTRs, can have regulatory functions related to, for example, translation rate and mRNA stability. Thus, these fragments of SDFs can be isolated for use as elements of gene constructs for regulated production of polynucleotides encoding desired polypeptides.

Introns of genomic DNA segments might also have regulatory functions. Sometimes regulatory elements, especially transcription enhancer or suppressor elements, are found within introns. Also, elements related to stability of heteronuclear RNA and efficiency of splicing and of transport to the cytoplasm for translation can be found in intron elements.

Thus, these segments can also find use as elements of expression vectors intended for use to transform plants.

Just as with promoters UTR sequences and intron/exon junctions can vary from those shown in Tables 1 and 2. Such changes from those sequences preferably will not affect the regulatory activity of the UTRs or intron/exon junction sequences on expression, transcription, or translation unless selected to do so. However, in some instances, down- or up-regulation of such activity may be desired to modulate traits or phenotypic or *in vitro* activity.

I.G. Coding Sequences

Isolated polynucleotides of the invention can include coding sequences that encode polypeptides comprising an amino acid sequence encoded by sequences in Tables 1 and 2 or an amino acid sequence presented in Tables 1 and 2.

A nucleotide sequence encodes a polypeptide if a cell (or a cell free *in vitro* system) expressing that nucleotide sequence produces a polypeptide having the recited amino acid sequence when the nucleotide sequence is transcribed and the primary transcript is subsequently processed and translated by a host cell (or a cell free *in vitro* system) harboring the nucleic acid. Thus, an isolated nucleic acid that encodes a particular amino acid sequence can be a genomic sequence comprising exons and introns or a cDNA sequence that represents the product of splicing thereof. An isolated nucleic acid encoding an amino acid sequence also encompasses heteronuclear RNA, which contains sequences that are spliced out during expression, and mRNA, which lacks those sequences.

Coding sequences can be constructed using chemical synthesis techniques or by isolating coding sequences or by modifying such synthesized or isolated coding sequences as described above.

In addition to coding sequences encoding the polypeptide sequences of Tables 1 and 2, which are native to corn, *Arabidopsis*, soybean, rice, wheat, and other plants the isolated polynucleotides can be polynucleotides that encode variants, fragments, and fusions of those native proteins. Such polypeptides are described below in part II.

In variant polynucleotides generally, the number of substitutions, deletions or insertions is preferably less than 20%, more preferably less than 15%; even more preferably less than 10%, 5%, 3% or 1% of the number of nucleotides comprising a particularly exemplified sequence. It is generally expected that non-degenerate nucleotide sequence changes that result in 1 to 10, more preferably 1 to 5 and most preferably 1 to 3 amino acid insertions, deletions or

substitutions will not greatly affect the function of an encoded polypeptide. The most preferred embodiments are those wherein 1 to 20, preferably 1 to 10, most preferably 1 to 5 nucleotides are added to, deleted from and/or substituted in the sequences specifically disclosed in Tables 1 and 2.

Insertions or deletions in polynucleotides intended to be used for encoding a polypeptide preferably preserve the reading frame. This consideration is not so important in instances when the polynucleotide is intended to be used as a hybridization probe.

II. Polypeptides and Proteins

IIA. Native polypeptides and proteins

Polypeptides within the scope of the invention include both native proteins as well as variants, fragments, and fusions thereof. Polypeptides of the invention are those encoded by any of the six reading frames of sequences shown in Tables 1 and 2, preferably encoded by the three frames reading in the 5' to 3' direction of the sequences as shown.

Native polypeptides include the proteins encoded by the sequences shown in Tables 1 and 2. Such native polypeptides include those encoded by allelic variants.

Polypeptide and protein variants will exhibit at least 75% sequence identity to those native polypeptides of Tables 1 and 2. More preferably, the polypeptide variants will exhibit at least 85% sequence identity; even more preferably, at least 90% sequence identity; more preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity. Fragments of polypeptide or fragments of polypeptides will exhibit similar percentages of sequence identity to the relevant fragments of the native polypeptide. Fusions will exhibit a similar percentage of sequence identity in that fragment of the fusion represented by the variant of the native peptide.

Furthermore, polypeptide variants will exhibit at least one of the functional properties of the native protein. Such properties include, without limitation, protein interaction, DNA interaction, biological activity, immunological activity, receptor binding, signal transduction, transcription activity, growth factor activity, secondary structure, three-dimensional structure, etc. As to properties related to *in vitro* or *in vivo* activities, the variants preferably exhibit at least 60% of the activity of the native protein; more preferably at least 70%, even more preferably at least 80%, 85%, 90% or 95% of at least one activity of the native protein.

One type of variant of native polypeptides comprises amino acid substitutions, deletions and/or insertions. Conservative substitutions are preferred to maintain the function or activity of the polypeptide.

Within the scope of percentage of sequence identity described above, a polypeptide of the invention may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide.

A.1 Antibodies

Isolated polypeptides can be utilized to produce antibodies. Polypeptides of the invention can generally be used, for example, as antigens for raising antibodies by known techniques. The resulting antibodies are useful as reagents for determining the distribution of the antigen protein within the tissues of a plant or within a cell of a plant. The antibodies are also useful for examining the production level of proteins in various tissues, for example in a wild-type plant or following genetic manipulation of a plant, by methods such as Western blotting.

Antibodies of the present invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the polypeptides of the invention are first used to immunize a suitable animal, such as a mouse, rat, rabbit, or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies as detection reagents. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization.

Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating the blood at 4°C for 2-18 hours. The serum is recovered by centrifugation (e.g., 1,000xg for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the method of Kohler and Milstein, *Nature* 256: 495 (1975), or modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen

cells can be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate, or well, coated with the protein antigen. B-cells producing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (e.g., hypoxanthine, aminopterin, thymidine medium, HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected Mab-secreting hybridomas are then cultured either *in vitro* (*e.g.*, in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

Other methods for sustaining antibody-producing B-cell clones, such as by EBV transformation, are known.

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TNB) to a blue pigment, quantifiable with a spectrophotometer.

A.2 In Vitro Applications of Polypeptides

Some polypeptides of the invention will have enzymatic activities that are useful *in vitro*. For example, the soybean trypsin inhibitor (Kunitz) family is one of the numerous families of proteinase inhibitors. It comprises plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases. Thus, these peptides find *in vitro* use in protein purification protocols and perhaps in therapeutic settings requiring topical application of protease inhibitors.

Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen and is also involved in chlorophyll biosynthesis(Kaczor et al. (1994) Plant Physiol. 1-4: 1411-7; Smith (1988) Biochem. J. 249: 423-8; Schneider (1976) Z. naturforsch. [C] 31: 55-63). Thus, ALAD proteins can be used as catalysts in synthesis of heme derivatives. Enzymes of biosynthetic pathways generally can be used as catalysts for *in vitro* synthesis of the compounds representing products of the pathway.

Polypeptides encoded by SDFs of the invention can be engineered to provide purification reagents to identify and purify additional polypeptides that bind to them. This allows one to identify proteins that function as multimers or elucidate signal transduction or metabolic pathways. In the case of DNA binding proteins, the polypeptide can be used in a similar manner to identify the DNA determinants of specific binding (S. Pierrou et al., *Anal. Biochem.* 229:99 (1995), S. Chusacultanachai et al., *J. Biol. Chem.* 274:23591 (1999), Q. Lin et al., *J. Biol. Chem.* 272:27274 (1997)).

II.B . POLYPEPTIDE VARIANTS , FRAGMENTS, AND FUSIONS

Generally, variants , fragments, or fusions of the polypeptides encoded by the maximum length sequence(MLS) can exhibit at least one of the activities of the identified domains and/or related polypeptides described in Sections (C) and (D) of Table 1 corresponding to the MLS of interest.

II.B .(1) Variants

A type of variant of the native polypeptides comprises amino acid substitutions. Conservative substitutions, described above (see II.), are preferred to maintain the function or activity of the polypeptide. Such substitutions include conservation of charge, polarity, hydrophobicity, size, etc. For example, one or more amino acid residues within the sequence can be substituted with another amino acid of similar polarity that acts as a functional equivalent, for example providing a hydrogen bond in an enzymatic catalysis. Substitutes for an amino acid within an exemplified sequence are preferably made among the members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Within the scope of percentage of sequence identity described above, a polypeptide of the invention may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide. Amino acid substitutions may also be made in the sequences; conservative substitutions being preferred.

One preferred class of variants are those that comprise (1) the domain of an encoded polypeptide and/or (2) residues conserved between the encoded polypeptide and related polypeptides. For this class of variants, the encoded polypeptide sequence is changed by insertion, deletion, or substitution at positions flanking the domain and/or conserved residues.

Another class of variants includes those that comprise an encoded polypeptide sequence that is changed in the domain or conserved residues by a conservative substitution.

Yet another class of variants includes those that lack one of the *in vitro* activities, or structural features of the encoded polypeptides. One example is polypeptides or proteins produced from genes comprising dominant negative mutations. Such a variant may comprise an encoded polypeptide sequence with non-conservative changes in a particular domain or group of conserved residues.

II.A.(2) FRAGMENTS

Fragments of particular interest are those that comprise a domain identified for a polypeptide encoded by an MLS of the instant invention and variants thereof. Also, fragments that comprise at least one region of residues conserved between an MLS encoded polypeptide and its related polypeptides are of great interest. Fragments are sometimes useful as polypeptides corresponding to genes comprising dominant negative mutations are.

II.A.(3)FUSIONS

Of interest are chimeras comprising (1) a fragment of the MLS encoded polypeptide or variants thereof of interest and (2) a fragment of a polypeptide comprising the same domain. For example, an AP2 helix encoded by a MLS of the invention fused to second AP2 helix from ANT protein, which comprises two AP2 helices. The present invention also encompasses fusions of MLS encoded polypeptides, variants, or fragments thereof fused with related proteins or fragments thereof.

DEFINITION OF DOMAINS

The polypeptides of the invention may possess identifying domains as shown in Table 1. Specific domains within the MLS encoded polypeptides are indicated in Table 1. In addition, the domains within the MLS encoded polypeptide can be defined by the region that exhibits at least 70% sequence identity with the consensus sequences listed in the detailed description below of each of the domains.

The majority of the protein domain descriptions given below are obtained from Prosite,
(http//www.expasy.ch/prosite/), and Pfam,
(http//pfam.wustl.edu/browse.shtml).

1. (AAA) AAA-protein family signature

A large family of ATPases has been described [1 to 5] whose key feature is that they share a conserved region of about 220 amino acids that contains anATP-binding site. This family is now called AAA, for 'A'TPases 'A'ssociated with diverse cellular 'A'ctivities. The proteins that belong to this family either contain one or two AAA domains. Proteins containing two AAA domains:

- Mammalian and drosophila NSF (N-ethylmaleimide-sensitive fusion protein) and the fungal homolog, SEC18. These proteins are involved in intracellular transport between the endoplasmic reticulum and Golgi, as well as between different Golgi cisternae.
- Mammalian transitional endoplasmic reticulum ATPase (previously known as p97 or VCP) which is involved in the transfer of membranes from the endoplasmic reticulum to the golgi apparatus. This protein forms a ring-shaped homooligomer composed of six subunits. The yeast homolog is CDC48 and it may play a role in spindle pole proliferation.
- Yeast protein PAS1, essential for peroxisome assembly and the related protein PAS1 from Pichia pastoris.
- Yeast protein AFG2.
- Sulfolobus acidocaldarius protein SAV and Halobacterium salinarium cdcH which may be part of a transduction pathway connecting light to cell division.

Proteins containing a single AAA domain:

- Escherichia coli and other bacteria ftsH (or hflB) protein. FtsH is an ATP-dependent zinc metallopeptidase that seems to degrade the heat-shock sigma-32 factor.
  It is an integral membrane protein with a large cytoplasmic C-terminal domain that contain both the AAA and the protease domains.
- Yeast protein YME1, a protein important for maintaining the integrity of the mitochondrial compartment. YME1 is also a zinc-dependent protease.
- Yeast protein AFG3 (or YTA10). This protein also seems to contain a AAA domain followed by a zinc-dependent protease domain.

Subunits from the regulatory complex of the 26S proteasome [6] which is involved in the ATP-dependent degradation of ubiquitinated proteins:

a) Mammalian subunit 4 and homologs in other higher eukaryotes, in yeast (gene YTA5) and fission yeast (gene mts2).

b) Mammalian subunit 6 (TBP7) and homologs in other higher eukaryotes and in yeast (gene YTA2).

c) Mammalian subunit 7 (MSS1) and homologs in other higher eukaryotes and in yeast (gene CIM5 or YTA3).

d) Mammalian subunit 8 (P45) and homologs in other higher eukaryotes and in yeast (SUG1 or CIM3 or TBY1) and fission yeast (gene let1).

Other probable subunits such as human TBP1 which seems to influences HIV gene expression by interacting with the virus tat transactivator protein and yeast YTA1 and YTA6.

- Yeast protein BCS1, a mitochondrial protein essential for the expression of the Rieske iron-sulfur protein.
- Yeast protein MSP1, a protein involved in intramitochondrial sorting of proteins.
- Yeast protein PAS8, and the corresponding proteins PAS5 from Pichia pastoris and PAY4 from Yarrowia lipolytica.
- Mouse protein SKD1 and its fission yeast homolog (SpAC2G11.06).
- Caenorhabditis elegans meiotic spindle formation protein mei-1.
- Yeast protein SAP1.
- Yeast protein YTA7.
- Mycobacterium leprae hypothetical protein A2126A.

It is proposed that, in general, the AAA domains in these proteins act as ATP-dependent protein clamps [5]. In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, there is a highly conserved region located in the central part of the domain which was used to develop a signature pattern.

Consensus pattern: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-x-R

[1] Froehlich K.-U., Fries H.W., Ruediger M., Erdmann R., Botstein D., Mecke D. J. Cell Biol. 114:443-453(1991).

[2] Erdmann R., Wiebel F.F., Flessau A., Rytka J., Beyer A., Froehlich K.-U., Kunau W.-H. Cell 64:499-510(1991).

[3] Peters J.-M., Walsh M.J., Franke W.W. EMBO J. 9:1757-1767(1990).

[4] Kunau W.-H., Beyer A., Goette K., Marzioch M., Saidowsky J., Skaletz-Rorowski A., Wiebel F.F. Biochimie 75:209-224(1993).

[5] Confalonieri F., Duguet M. BioEssays 17:639-650(1995).[6] Hilt W., Wolf D.H. Trends Biochem. Sci. 21:96-102(1996).

2. ABC Membrane (ABC transporter transmembrane region). This family represents a unit of six transmembrane helices. Many members of the ABC transporter family (ABC_tran)have two such regions. See also descriptions of ABC Tran, below, and ABC2 membrane, above.

3. (ABC Tran) ABC transporters family signature. On the basis of sequence similarities a family of related ATP-bindingproteins has been characterized [1 to 5]. These proteins are associated with avariety of distinct biological processes in both prokaryotes and eukaryotes, but a majority of them are involved in active transport of small hydrophilic molecules across the cytoplasmic membrane. All these proteins share a conserved domain of some two hundred amino acid residues, which includes an ATP-binding site. These proteins are collectively known as ABC transporters. Proteins known to belong to this family are listed below (references are only provided for recently determined sequences).In prokaryotes: - Active transport systems components: alkylphosphonate uptake(phnC/phnK/ phnL); arabinose (araG); arginine (artP); dipeptide (dciAD;dppD/dppF); ferric enterobactin (fepC); ferrichrome (fhuC); galactoside (mglA); glutamine (glnQ); glycerol-3-phosphate (ugpC); glycine betaine/L-proline (proV); glutamate/aspatate (gltL); histidine (hisP); iron(III) (sfuC), iron(III) dicitrate (fecE); lactose (lacK); leucine/isoleucine/valine (braF/braG;livF/livG); maltose (malK); molybdenum (modC); nickel (nikD/ nikE); oligopeptide (amiE/amiF;oppD/oppF); peptide (sapD/sapF); phosphate (pstB); putrescine (potG); ribose (rbsA); spermidine/putrescine (potA); sulfate (cysA); vitamin B12 (btuD). - Hemolysin/leukotoxin export proteins hlyB, cyaB and lktB. - Colicin V export protein cvaB. - Lactococcin export protein lcnC [6]. - Lantibiotic transport proteins nisT (nisin) and spaT (subtilin). - Extracellular proteases B and C export protein prtD. - Alkaline protease secretion protein aprD. - Beta-(1,2)-glucan export proteins chvA and ndvA. - Haemophilus influenzae capsule-polysaccharide export protein bexA. - Cytochrome c biogenesis proteins ccmA (also known as cycV and helA). - Polysialic acid transport protein kpsT. - Cell division associated ftsE protein (function unknown). - Copper processing protein nosF from Pseudomonas

stutzeri. - Nodulation protein nodI from Rhizobium (function unknown). - Escherichia coli proteins cydC and cydD. - Subunit A of the ABC excision nuclease (gene uvrA). - Erythromycin resistance protein from Staphylococcus epidermidis (gene msrA). - Tylosin resistance protein from Streptomyces fradiae (gene tlrC) [7]. - Heterocyst differentiation protein (gene hetA) from Anabaena PCC 7120. - Protein P29 from Mycoplasma hyorhinis, a probable component of a high affinity transport system. - yhbG, a putative protein whose gene is linked with ntrA in many bacteria such as Escherichia coli, Klebsiella pneumoniae, Pseudomonas putida, Rhizobium meliloti and Thiobacillus ferrooxidans. - Escherichia coli and related bacteria hypothetical proteins yabJ, yadG, yagC, ybbA, ycjW, yddA, yehX, yejF, yheS, yhiG, yhiH, yjcW, yjjK, yojI, yrbF and ytfR.In eukaryotes: - The multidrug transporters (Mdr) (P-glycoprotein), a family of closely related proteins which extrude a wide variety of drugs out of the cell (for a review see [8]). - Cystic fibrosis transmembrane conductance regulator (CFTR), which is most probably involved in the transport of chloride ions. - Antigen peptide transporters 1 (TAP1, PSF1, RING4, HAM-1, mtp1) and 2 (TAP2, PSF2, RING11, HAM-2, mtp2), which are involved in the transport of antigens from the cytoplasm to a membrane-bound compartment for association with MHC class I molecules. - 70 Kd peroxisomal membrane protein (PMP70). - ALDP, a peroxisomal protein involved in X-linked adrenoleukodystrophy [9]. - Sulfonylurea receptor [10], a putative subunit of the B-cell ATP-sensitive potassium channel. - Drosophila proteins white (w) and brown (bw), which are involved in the import of ommatidium screening pigments. - Fungal elongation factor 3 (EF-3). - Yeast STE6 which is responsible for the export of the a-factor pheromone. - Yeast mitochondrial transporter ATM1. - Yeast MDL1 and MDL2. - Yeast SNQ2. - Yeast sporidesmin resistance protein (gene PDR5 or STS1 or YDR1). - Fission yeast heavy metal tolerance protein hmt1. This protein is probably involved in the transport of metal-bound phytochelatins. - Fission yeast brefeldin A resistance protein (gene bfr1 or hba2). - Fission yeast leptomycin B resistance protein (gene pmd1). - mbpX, a hypothetical chloroplast protein from Liverwort. - Prestalk-specific protein tagB from slime mold. This protein consists of two domains: a N-terminal subtilase catalytic domain and a C-terminal ABC transporter domain.As a signature pattern for this class of proteins, a conserved region which is located between the 'A' and the 'B' motifs of the ATP-binding site was used.

Consensus pattern: [LIVMFYC]-[SA]-[SAPGLVFYKQH]-G-[DENQMW]-[KRQASPCLIMFW]-[KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]-{PHY}-[LIVMFW]-[SAGCLIVP]-{FYWHP}-{KRHP}-[LIVMFYWSTA] The ATP-binding region is

duplicated in araG, mdl, msrA, rbsA, tlrC, uvrA, yejF, Mdr's, CFTR, pmd1 and in EF-3. In some of those proteins, the above pattern only detect one of the two copies of the domain. The proteins belonging to this family also contain one or two copies of the ATP-binding motifs 'A' and 'B'.

[1] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).

[2] Higgins C.F., Gallagher M.P., Mimmack M.M., Pearce S.R. BioEssays 8:111-116(1988).

[3] Higgins C.F., Hiles I.D., Salmond G.P.C., Gill D.R., Downie J.A., Evans I.J., Holland I.B., Gray L., Buckels S.D., Bell A.W., Hermodson M.A. Nature 323:448-450(1986).

[4] Doolittle R.F., Johnson M.S., Husain I., van Houten B., Thomas D.C., Sancar A. Nature 323:451-453(1986).

[5] Blight M.A., Holland I.B. Mol. Microbiol. 4:873-880(1990).

[6] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L. Appl. Environ. Microbiol. 58:1952-1961(1992).

[7] Rosteck P.R. Jr., Reynolds P.A., Hershberger C.L. Gene 102:27-32(1991).

[8] Gottesman M.M., Pastan I. J. Biol. Chem. 263:12163-12166(1988).

[9] Valle D., Gaertner J. Nature 361:682-683(1993).

[10] Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV, Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J., Nelson D.A. Science 268:423-426(1995).

4. (ACBP)

Acyl-CoA-binding protein signature

Acyl-CoA-binding protein (ACBP) is a small (10 Kd) protein that binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters [1]. ACBP is also known as diazepam binding inhibitor (DBI) or endozepine (EP) because of its ability to displace diazepam from the benzodiazepine (BZD) recognition site located on the GABA type A receptor. It is therefore possible that this protein also acts as a neuropeptide to modulate the action of the GABA receptor [2].ACBP is a highly conserved protein of about 90 residues that has been so far found in vertebrates, insects and yeast. ACBP is also related to the N-terminal section of a probable transmembrane protein of

unknown function whichhas been found in mammals. As a signature pattern, the region that corresponds to residues 19 to 37 in mammalian ACBP was selected.

Consensus pattern: P-[STA]-x-[DEN]-x-[LIVMF]-x(2)-[LIVMFY]-Y-[GSTA]-x-[FY]-K- Q-[STA](2)-x-G-

[1] Rose T.M., Schultz E.R., Todaro G.J. Proc. Natl. Acad. Sci. U.S.A. 89:11287-11291(1992).
[2] Costa E., Guidotti A. Life Sci. 49:325-344(1991).

5. (AIRS)
AIR synthase related proteins

This family includes Hydrogen expression/formation protein HypE, AIR synthases, FGAM synthase and selenide, water dikinase.

6. (AMP-binding)
Putative AMP-binding domain signature

It has been shown [1 to 5] that a number of prokaryotic and eukaryotic enzymes which all probably act via an ATP-dependent covalent binding of AMP to their substrate, share a region of sequence similarity. These enzymes are: - Insects luciferase (luciferin 4-monooxygenase). Luciferase produces light by catalyzing the oxidation of luciferin in presence of ATP and molecular oxygen. - Alpha-aminoadipate reductase from yeast (gene LYS2). This enzyme catalyzes the activation of alpha-aminoadipate by ATP-dependent adenylation and the reduction of activated alpha-aminoadipate by NADPH. - Acetate--CoA ligase (acetyl-CoA synthetase), an enzyme that catalyzes the formation of acetyl-CoA from acetate and CoA. - Long-chain-fatty-acid--CoA ligase, an enzyme that activates long-chain fatty acids for both the synthesis of cellular lipids and their degradation via beta-oxidation. - 4-coumarate--CoA ligase (4CL), a plant enzyme that catalyzes the formation of 4-coumarate-CoA from 4-coumarate and coenzyme A; the branchpoint reactions between general phenylpropanoid metabolism and pathways leading to various specific end products. -

O-succinylbenzoic acid--CoA ligase (OSB-CoA synthetase) (gene menE) [6], a bacterial enzyme involved in the biosynthesis of menaquinone (vitamin K2). - 4-Chlorobenzoate--CoA ligase (EC 6.2.1.-) (4-CBA--CoA ligase) [7], a Pseudomonas enzyme involved in the degradation of 4-CBA. - Indoleacetate--lysine ligase (IAA-lysine synthetase) [8], an enzyme from Pseudomonas syringae that converts indoleacetate to IAA-lysine. - Bile acid-CoA ligase (gene baiB) from Eubacterium strain VPI 12708 [4]. This enzyme catalyzes the ATP-dependent formation of a variety of C-24 bile acid-CoA. - Crotonobetaine/carnitine-CoA ligase (EC 6.3.2.-) from Escherichia coli (gene caiC). - L-(alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase (ACV synthetase) from various fungi (gene acvA or pcbAB). This enzyme catalyzes the first step in the biosynthesis of penicillin and cephalosporin, the formation of ACV from the constituent amino acids. The amino acids seem to be activated by adenylation. It is a protein of around 3700 amino acids that contains three related domains of about 1000 amino acids. - Gramicidin S synthetase I (gene grsA) from Bacillus brevis. This enzyme catalyzes the first step in the biosynthesis of the cyclic antibiotic gramicidin S, the ATP-dependent racemization of phenylalanine - Tyrocidine synthetase I (gene tycA) from Bacillus brevis. The reaction carried out by tycA is identical to that catalyzed by grsA - Gramicidin S synthetase II (gene grsB) from Bacillus brevis. This enzyme is a multifunctional protein that activates and polymerizes proline, valine, ornithine and leucine. GrsB consists of four related domains. - Enterobactin synthetase components E (gene entE) and F (gene entF) from Escherichia coli. These two enzymes are involved in the ATP-dependent activation of respectively 2,3-dihydroxybenzoate and serine during enterobactin (enterochelin) biosynthesis. - Cyclic peptide antibiotic surfactin synthase subunits 1, 2 and 3 from Bacillus subtilis. Subunits 1 and 2 contains three related domains while subunit 3 only contains a single domain. - HC-toxin synthetase (gene HTS1) from Cochliobolus carbonum. This enzyme activates the four amino acids (Pro, L-Ala, D-Ala and 2-amino-9,10-epoxi-8-oxodecanoic acid) that make up HC-toxin, a cyclic tetrapeptide. HTS1 consists of four related domains.There are also some proteins, whose exact function is not yet known, but whichare, very probably, also AMP-binding enzymes. These proteins are: - ORA (octapeptide-repeat antigen), a Plasmodium falciparum protein whose function is not known but which shows a high degree of similarity with the above proteins. - AngR, a Vibrio anguillarum protein. AngR is thought to be a transcriptional activator which modulates the anguibactin (an iron-binding siderophore) biosynthesis gene cluster operon. But it is believed [9], that angR is not a DNA-binding protein, but rather an enzyme involved in the biosynthesis of anguibactin. This conclusion is based on three facts: the presence of the AMP-binding domain; the size of

angR (1048 residues), which is far bigger than any bacterial transcriptional protein; and the presence of a probable S-acyl thioesterase immediately downstream of angR. - A hypothetical protein in mmsB 3'region in Pseudomonas aeruginosa. - Escherichia coli hypothetical protein ydiD. - Yeast hypothetical protein YBR041w. - Yeast hypothetical protein YBR222c. - Yeast hypothetical protein YER147c.All these proteins contain a highly conserved region very rich in glycine, serine, and threonine which is followed by a conserved lysine. A parallel can be drawn between this type of domain and the G-x(4)-G-K-[ST] ATP-/GTP-binding 'P-loop' domain or the protein kinases G-x-G-x(2)-[SG]-x(10,20)-KATP-binding domains.

Consensus pattern: [LIVMFY]-x(2)-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR] In a majority of cases the residue that follows the Lys at the end of the pattern is a Gly.

[1] Toh H. Protein Seq. Data Anal. 4:111-117(1991).
[2] Smith D.J., Earl A.J., Turner G. EMBO J. 9:2743-2750(1990).
[3] Schroeder J. Nucleic Acids Res. 17:460-460(1989).
[4] Mallonee D.H., Adams J.L., Hylemon P.B. J. Bacteriol. 174:2065-2071(1992).
[5] Turgay K., Krause M., Marahiel M.A. Mol. Microbiol. 6:529-546(1992).
[6] Driscoll J.R., Taber H.W. J. Bacteriol. 174:5063-5071(1992).
[7] Babbitt P.C., Kenyon G.L., Matin B.M., Charest H., Sylvestre M., Scholten J.D., Chang K.-H., Liang P.-H., Dunaway-Mariano D. Biochemistry 31:5594-5604(1992).
[8] Farrell D.H., Mikesell P., Actis L.A., Crosa J.H. Gene 86:45-51(1990).

7. AP2 domain

This 60 amino acid residue domain can bind to DNA [1]. This domain is plant specific. Members of this family are suggested to be related to pyridoxal phosphate-binding domains such as found in aminotran_2 [3]. AP2 domains are also described in Jofuku et al., co-pending U.S. Patent applications 08/700,152, 08/879,827, 08/912,272, 09/026,039.

[1] Ohme-takagi M, Shinshi H; Plant Cell 1995;7:173-182.
[2] Weigel D; Plant Cell 1995;7:388-389.
[3] Mushegian AR, Koonin EV; Genetics 1996;144:817-828.

8. ARID

The ARID domain is an AT-Rich Interaction domain sharing structural homology to DNA replication and repair nucleases and polymerases.

[1] Herrscher RF, Kaplan MH, Lelsz DL, Das C, Scheuermann R, Tucker PW; Genes Dev 1995;9:3067-3082.

[2] Yuan YC, Whitson RH, Liu Q, Itakura K, Chen Y; Nat Struct Biol 1998;5:959-964.

9. (ATP synt)

ATP synthase gamma subunit signature

ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a componentof the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), and a catalytic core, called coupling factor CF(1). The former acts as a proton channel; the latter is composed of five subunits, alpha, beta, gamma, delta and epsilon. Subunit gamma is believed to be important in regulating ATPase activity and the flow of protons through the CF(0) complex. The best conserved region of the gamma subunit [3] is its C-terminus which seems to be essential for assembly and catalysis. As a signature pattern to detect ATPase gamma subunits, a14 residue conserved segment where the last amino acid is found one to three residues from the C-terminal extremity was used.

Consensus pattern: [IV]-T-x-E-x(2)-[DE]-x(3)-G-A-x-[SAKR]- Note: Pea chloroplast gamma and two Bacillus species gamma subunits are not detected by this motif.

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).

[2] Senior A.E. Physiol. Rev. 68:177-231(1988).

[3] Miki J., Maeda M., Mukohata Y., Futai M. FEBS Lett. 232:221-226(1988).

10. (ATP Synt A)

Synthase a subunit signature

ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria,and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1).The CF(0) a subunit, also called protein 6, is a key component of the proton channel; it may play a direct role in translocating protons across the membrane. It is a highly hydrophobic protein that has been predicted to contain 8 transmembrane regions [3].Sequence comparison of a subunits from all available sources reveals very few conserved regions. The best conserved region is located in what is predicted to be the fifth transmembrane domain. This region contains three perfectly conserved residues: an arginine, a leucine and an asparagine. Mutagenesis experiments of ATPase activity. This region was selected as a signature pattern.

Consensus pattern: [STAGN]-x-[STAG]-[LIVMF]-R-L-x-[SAGV]-N-[LIVMT] [R is important for proton translocation]

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).
[2] Senior A.E. Physiol. Rev. 68:177-231(1988).
[3] Lewis M.L., Chang J.A., Simoni R.D. J. Biol. Chem. 265:10541-10550(1990).
[4] Cain B.D., Simoni R.D. J. Biol. Chem. 264:3292-3300(1989).

11. ATP synthase B

Part of the CF(0) (base unit) of the ATP synthase. The base unit is thought to translocate protons through membrane (inner membrane in mitochondria, thylakoid membrane in plants, cytoplasmic membrane in bacteria). The B subunits are thought to interact with the stalk of the CF(1) subunits.

12. (ATP synt C)

ATP synthase c subunit signature

ATP synthase (proton-translocating ATPase) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria,and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1).The CF(0) c subunit (also called protein 9, proteolipid, or subunit III) [3,4]is a highly hydrophobic protein of about 8 Kd which has been implicated in the proton-conducting activity of ATPase. Structurally subunit c consist of two long terminal hydrophobic regions, which probably span the membrane, and a central hydrophilic region. N,N'-dicyclohexylcarbodiimide (DCCD) can bind covalently to subunit c and thereby abolish the ATPase activity. DCCD binds to a specific glutamate or aspartate residue which is located in the middle ofthe second hydrophobic region near the C-terminus of the protein. A signature pattern which includes the DCCD-binding residue was derived.

Consensus pattern: [GSTA]-R-[NQ]-P-x(10)-[LIVMFYW](2)-x(3)-[LIVMFYW]-x-[DE] [D or E binds DCCD]

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).
[2] Senior A.E. Physiol. Rev. 68:177-231(1988).
[3] Ivaschenko A.T., Karpenyuk T.A., Ponomarenko S.V. Biokhimiia 56:406-419(1991).
[4] Recipon H., Perasso R., Adoutte A., Quetier F. J. Mol. Evol. 34:292-303(1992).

13. (ATP synt DE)
ATP synthase, Delta/Epsilon chain

Part of the ATP synthase CF(1). These subunits are part of the head unit of the ATP synthase. The subunits are called delta and epsilon in human and metozoan species but in bacterial species the delta (D) subunit is theequivalent to the Oligomycin sensitive subunit (OSCP) in metozoans.

14. (ATP synt ab)
ATP synthase alpha and beta subunits signature

ATP synthase (proton-translocating ATPase) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria,and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), and a catalytic core, called coupling factor CF(1). The former acts as a proton channel; the latter is composed of five subunits, alpha, beta, gamma, delta and epsilon. The sequences of subunits alpha and beta are related and both contain a nucleotide-binding site for ATP and ADP. The beta chain has catalytic activity, while the alpha chain is a regulatory subunit. Vacuolar ATPases [3] (V-ATPases) are responsible for acidifying a variety of intracellular compartments in eukaryotic cells. Like F-ATPases, they are oligomeric complexes of a transmembrane and a catalytic sector. The sequenceof the largest subunit of the catalytic sector (70 Kd) is related to that ofF-ATPase beta subunit, while a 60 Kd subunit, from the same sector, is related to the F-ATPases alpha subunit [4].Archaebacterial membrane-associated ATPases are composed of three subunits.The alpha chain is related to F-ATPases beta chain and the beta chain is related to F-ATPases alpha chain [4].A protein highly similar to F-ATPase beta subunits is found [5] in some bacterial apparatus involved in a specialized protein export pathway that proceeds without signal peptide cleavage. This protein is known as fliI in Bacillus and Salmonella, Spa47 (mxiB) in Shigella flexneri, HrpB6 in Xanthomonas campestris and yscN in Yersinia virulence plasmids.To detect these ATPase subunits, a segment of ten amino-acid residues, containing two conserved serines, as a signature pattern was selected. The first serine seems to be important for catalysis - in the ATPase alpha chain at least - as its mutagenesis causes catalytic impairment.

Consensus pattern: P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S [The first S is a putative active site residue]

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).

[2] Senior A.E. Physiol. Rev. 68:177-231(1988).

[3] Nelson N. J. Bioenerg. Biomembr. 21:553-571(1989).

[4] Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., Manolson M.F., Poole R.J., Date T., Oshima T., Konishi J., Denda K., Yoshida M. Proc. Natl. Acad. Sci. U.S.A. 86:6661-665(1989).

[5] Dreyfus G., Williams A.W., Kawagishi I., MacNab R.M. J. Bacteriol. 175:3131-3138(1993).

15. (ATP synt ab C)

ATP synthase ab C terminal.

Number of members: 190

[1] Abrahams JP, Leslie AG, Lutter R, Walker JE; Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria." Nature 1994;370:621-628.

16. (A deaminase)

Adenosine and AMP deaminase signature

Adenosine deaminase catalyzes the hydrolytic deamination ofadenosine into inosine. AMP deaminase catalyzes the hydrolytic deamination of AMP into IMP. It has been shown [1] that these two types of enzymes share three regions of sequence similarities; these regions are centered on residues which are proposed to play an important role in the catalytic mechanism of these two enzymes. One of these regions, containing two conserved aspartic acid residues that are potential active site residues was selected.

Consensus pattern: [SA]-[LIVM]-[NGS]-[STA]-D-D-P [The two D's are putative active site residues]

[1] Chang Z., Nygaard P., Chinault A.C., Kellems R.E. Biochemistry 30:2273-2280(1991).

17. (Acetyltransf)

Acetyltransferase (GNAT) family.

This family contains proteins with N-acetyltransferase functions.

[1] Neuwald AF, Landsman D; Trends Biochem Sci 1997;22:154-155.

18. (Aconitase C)

Aconitase family signature

Aconitase (aconitate hydratase) (EC 4.2.1.3) [1] is the enzyme from the tricarboxylic acid cycle that catalyzes the reversible isomerization of citrate and isocitrate. Cis-aconitate is formed as an intermediary product during the course of the reaction. In eukaryotes two isozymes of aconitase are known to exist: one found in the mitochondrial matrix and the other found in the cytoplasm. Aconitase, in its active form, contains a 4Fe-4S iron-sulfur cluster; three cysteine residues have been shown to be ligands of the 4Fe-4S cluster.It has been shown that the aconitase family also contains the followingproteins: - Iron-responsive element binding protein (IRE-BP). IRE-BP is a cytosolic protein that binds to iron-responsive elements (IREs). IREs are stem-loop structures found in the 5'UTR of ferritin, and delta aminolevulinic acid synthase mRNAs, and in the 3'UTR of transferrin receptor mRNA. IRE-BP also express aconitase activity. - 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate isomerase), the enzyme that catalyzes the second step in the biosynthesis of leucine. - Homoaconitase (EC 4.2.1.36) (homoaconitate hydratase), an enzyme that participates in the alpha-aminoadipate pathway of lysine biosynthesis and that converts cis-homoaconitate into homoisocitric acid. - Esherichia coli protein ybhJ.As a signature for proteins from the aconitase family, two conserved regions that contain the three cysteine ligands of the 4Fe-4Scluster were selected.

Consensus pattern: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-[LIVMA] [C binds the iron-sulfur center]

Consensus pattern: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA] [The two C's bind the iron-sulfur center]

[1] Gruer M.J., Artymiuk P.J., Guest J.R. Trends Biochem. Sci. 22:3-6(1997).

19. (Acyl-CoA dh)

Acyl-CoA dehydrogenases signatures

Acyl-CoA dehydrogenases [1,2,3] are enzymes that catalyze the alpha, beta-dehydrogenation of acyl-CoA esters and transfer electrons to ETF, the electron transfer protein. Acyl-CoA dehydrogenases are FAD flavoproteins. This family currently includes: - Five eukaryotic isozymes that catalyze the first step of the beta-oxidation cycles for fatty acids with various chain lengths. These are short (SCAD) (EC 1.3.99.2), medium (MCAD) (EC 1.3.99.3), long (LCAD) (EC 1.3.99.13), very-long (VLCAD) and short/branched (SBCAD) chain acyl-CoA dehydrogenases. These enzymes are located in the mitochondrion. They are all homotetrameric proteins of about 400 amino acid residues except VLCAD which is a dimer and which contains, in its mature form, about 600 residues. - Glutaryl-CoA dehydrogenase (EC 1.3.99.7) (GCDH), which is involved in the catabolism of lysine, hydroxylysine and tryptophan. - Isovaleryl-CoA dehydrogenase (EC 1.3.99.10) (IVD), involved in the catabolism of leucine. - Acyl-coA dehydrogenases acsA and mmgC from Bacillus subtilis. - Butyryl-CoA dehydrogenase (EC 1.3.99.2) from Clostridium acetobutylicum. - Escherichia coli protein caiA [4]. - Escherichia coli protein aidB. Two conserved regions were selected as signature patterns. The first is located in the center of these enzymes, the second in the C-terminal section.

Consensus pattern: [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA]

Consensus pattern: [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)- [DEN]

[1] Tanaka K., Ikeda, Matsubara Y., Hyman D.B. Enzyme 38:91-107(1987).
[2] Matsubara Y., Indo Y., Naito E., Ozasa H., Glassberg R., Vockley J., Ikeda Y., Kraus J., Tanaka K. J. Biol. Chem. 264:16321-16331(1989).
[3] Aoyama T., Ueno I., Kamijo T., Hashimoto T. J. Biol. Chem. 269:19088-19094(1994).
[4] Eichler K., Bourgis F., Buchet A., Kleber H.-P., Mandrand-Berthelot M.-A. Mol. Microbiol. 13:775-786(1994).

20. (Acyl transf)
Acyl transferase domain

Number of members: 161

[1] Serre L, Verbree EC, Dauter Z, Stuitje AR, Derewenda ZS; Medline: 95286570 The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component." J Biol Chem 1995;270:12961-12964.

21. Acylphosphatase signatures

Acylphosphatase (EC 3.6.1.7) [1,2] catalyzes the hydrolysis of various acylphosphate carboxyl-phosphate bonds such as carbamyl phosphate, succinylphosphate, 1,3-diphosphoglycerate, etc. The physiological role of this enzymeis not yet clear. Acylphosphatase is a small protein of around 100 amino-acid residues. There are two known isozymes. One seems to be specific to muscular tissues, the other, called 'organ-common type', is found in many different tissues.While acylphosphatase have been so far only characterized in vertebrates,there are a number of bacterial and archebacterial hypothetical proteins that are highly similar to that enzyme and that probably possess the same activity.These proteins are: - Escherichia coli hypothetical protein yccX. - Bacillus subtilis hypothetical protein yflL. - Archaeoglobus fulgidus hypothetical protein AF0818. Two conserved regions were selected as signature patterns. The first is located in the N-terminal section, while the second is found in the central part ofthe protein sequence.

Consensus pattern: [LIV]-x-G-x-V-Q-G-V-x-[FM]-R

Consensus pattern: G-[FYW]-[AVC]-[KRQAM]-N-x(3)-G-x-V-x(5)-G

[1] Stefani M., Ramponi G. Life Chem. Rep. 12:271-301(1995).
[2] Stefani M., Taddei N., Ramponi G. Cell. Mol. Life Sci. 53:141-151(1997).

22. (Adap comp sub)
Clathrin adaptor complexes medium chain signatures.

Clathrin coated vesicles (CCV) mediate intracellular membrane traffic such asreceptor mediated endocytosis. In addition to clathrin, the CCV are composed of a number of other components including oligomeric complexes which are knownas adaptor or clathrin assembly

proteins (AP) complexes [1]. The adaptor complexes are believed to interact with the cytoplasmic tails of membrane proteins, leading to their selection and concentration. In mammals two type of adaptor complexes are known: AP-1 which is associated with the Golgi complex and AP-2 which is associated with the plasma membrane. Both AP-1 and AP-2 are heterotetramers that consist of two large chains - the adaptins - (gamma and beta' in AP-1; alpha and beta in AP-2); a medium chain (AP47 in AP-1; AP50 inAP-2) and a small chain (AP19 in AP-1; AP17 in AP-2). The medium chains of AP-1 and AP-2 are evolutionary related proteins of about 50 Kd. Homologs of AP47 and AP50 have also been found in Caenorhabditis elegans (genes unc-101 and ap50) [2] and yeast (gene APM1 or YAP54) [3].Some more divergent, but clearly evolutionary related proteins have also been found in yeast: APM2 and YBR288c., Two conserved regions were selected as signature patterns, one located in the N-terminal region, the other from the central section of these proteins.

Consensus pattern: [IVT]-[GSP]-W-R-x(2,3)-[GAD]-x(2)-[HY]-x(2)-N-x- [LIVMAFY](3)-D-[LIVM]-[LIVMT]-E

Consensus pattern: [LIV]-x-F-I-P-P-x-G-x-[LIVMFY]-x-L-x(2)-Y

[1] Pearse B.M., Robinson M.S. Annu. Rev. Cell Biol. 6:151-171(1990).
[2] Lee J., Jongeward G.D., Sternberg P.W. Genes Dev. 8:60-73(1994).
[3] Nakayama Y., Goebl M., O'Brine G.B., Lemmon S., Pingchang C.E., Kirchhausen T. Eur. J. Biochem. 202:569-574(1991).

23. (Adenylsucc synt)
Adenylosuccinate synthetase signatures

Adenylosuccinate synthetase (EC 6.3.4.4) [1] plays an important role in purinebiosynthesis, by catalyzing the GTP-dependent conversion of IMP and aspartic acid to AMP. Adenylosuccinate synthetase has been characterized from various sources ranging from Escherichia coli (gene purA) to vertebrate tissues. Invertebrates, two isozymes are present - one involved in purine biosynthesis and the other in the purine nucleotide cycle. Two conserved regions were selected as signature patterns. The first one is a perfectly conserved

octapeptide located in the N-terminal section and which is involved in GTP-binding [2]. The second one includes a lysine residue known [2] to be essential for the enzyme's activity.

Consensus pattern: Q-W-G-D-E-G-K-G

Consensus pattern: G-I-[GR]-P-x-Y-x(2)-K-x(2)-R [K is the active site residue]

[1] Wiesmueller L., Wittbrodt J., Noegel A.A., Schleicher M. J. Biol. Chem. 266:2480-2485(1991).

[2] Silva M.M., Poland B.W., Hoffman C.R., Fromm H.J., Honzatko R.B. J. Mol. Biol. 254:431-446(1995).

[3] Bouyoub A., Barbier G., Forterre P., Labedan B. 2.3.CO;2-"J. Mol. Biol. 261:144-154(1996).

24. (AdoHcyase)

S-adenosyl-L-homocysteine hydrolase signatures

S-adenosyl-L-homocysteine hydrolase (EC 3.3.1.1) (AdoHcyase) is an enzyme of the activated methyl cycle, responsible for the reversible hydratation of S-adenosyl-L-homocysteine into adenosine and homocysteine. AdoHcyase is anubiquitous enzyme which binds and requires NAD+ as a cofactor. AdoHcyase is a highly conserved protein [1] of about 430 to 470 amino acids. Two highly conserved regions were selected as signature patterns. The first pattern is located in the N-terminal section; the second is derived from aglycine-rich region in the central part of AdoHcyase; a region thought to be involved in NAD-binding.

Consensus pattern: [GSA]-[CS]-N-x-[FYLM]-S-[ST]-[QA]-[DEN]-x-[AV]-[AT]-[AD]-[AC]-[LIVMCG]

Consensus pattern: [GA]-[KS]-x(3)-[LIV]-x-G-[FY]-G-x-[VC]-G-[KRL]-G-x-[ASC]

[1] Sganga M.W., Aksamit R.R., Cantoni G.L., Bauer C.E. Proc. Natl. Acad. Sci. U.S.A. 89:6328-6332(1992).

25. AhpC/TSA family

This family contains proteins related to alkyl hydroperoxide reductaseComment: (AhpC) and thiol specific antioxidant (TSA).

[1] Chae HZ, Robison K, Poole LB, Church G, Storz G, Rhee SG, Proc Natl Acad Sci U S A 1994;91:7017-7021

26. (Aldose epim)

Aldose 1-epimerase putative active site Aldose 1-epimerase (EC 5.1.3.3) (mutarotase) is the enzyme responsible for the anomeric interconversion of D-glucose and other aldoses between their alpha- and beta-forms. The sequence of mutarotase from two bacteria, Acinetobacter calcoaceticus and Streptococcus thermophilus is available [1]. It has also been shown that, on the basis of extensive sequence similarities, a mutarotase domain seem to be present in the C-terminal half of the fungal GAL10 protein which encodes, in the N-terminal part, for UDP-glucose 4-epimerase. The best conserved region in the sequence of mutarotase is centered around a conserved histidine residue which may be involved in the catalytic mechanism.

Consensus pattern: [NS]-x-T-N-H-x-Y-[FW]-N-[LI]

[1] Poolman B., Royer T.J., Mainzer S.E., Schmidt B.F. J. Bacteriol. 172:4037-4047(1990).

27. (AlkA DNA repair)

Alkylbase DNA glycosidases alkA family signature

Alkylbase DNA glycosidases [1] are DNA repair enzymes that hydrolyzes the deoxyribose N-glycosidic bond to excise various alkylated bases from a damaged DNA polymer. In Escherichia coli there are two alkylbase DNA glycosidases: one (gene tag)which is constitutively expressed and which is specific for the removal of 3-methyladenine (EC 3.2.2.20), and one (gene alkA) which is induced during adaptation to alkylation and which

can remove a variety of alkylation products (EC 3.2.2.21). Tag and alkA do not share any region of sequence similarity. In yeast there is an alkylbase DNA glycosidase (gene MAG1) [2,3], which can remove 3-methyladenine or 7-methyladenine and which is structurally related to alkA. MAG and alkA are both proteins of about 300 amino acid residues. While the C- and N-terminal ends appear to be unrelated, there is a central region of about 130 residues which is well conserved. A portion of this region has been selected as a signature pattern .

Consensus pattern: G-I-G-x-W-[ST]-[AV]-x-[LIVMFY](2)-x-[LIVM]-x(8)-[MF]-x(2)-[ED]-D

[1] Lindahl T., Sedgwick B. Annu. Rev. Biochem. 57:133-157(1988).
[2] Berdal K.G., Bjoras M., Bjelland S., Seeberg E.C. EMBO J. 9:4563-4568(1990).
[3] Chen J., Derfler B., Samson L. EMBO J. 9:4569-4575(1990).

28. Ammonium transporters signature

A number of proteins involved in the transport of ammonium ions across amembrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are: - Yeast ammonium transporters MEP1, MEP2 and MEP3. - Arabidopsis thaliana high affinity ammonium transporter (gene AMT1). - Corynebacterium glutamicum ammonium and methylammonium transport system. - Escherichia coli putative ammonium transporter amtB. - Bacillus subtilis nrgA. - Mycobacterium tuberculosis hypothetical protein MtCY338.09c. - Synechocystis strain PCC 6803 hypothetical proteins sll0108, sll0537 and sll1017. - Methanococcus jannaschii hypothetical proteins MJ0058 and MJ1343. - Caenorhabditis elegans hypothetical proteins C05E11.4, F49E11.3 and M195.3. As expected by their transport function, these proteins are highly hydrophobic and seem to contain from 10 to 12 transmembrane domains. The best conserved region seems to be located in the fifth (or sixth) transmembrane region and is used as a signature pattern.

Consensus pattern: D-[FYWS]-A-G-[GSC]-x(2)-[IV]-x(3)-[SAG](2)-x(2)-[SAG]- [LIVMF]-x(3)-[LIVMFYWA](2)-x-[GK]-x-R

[1] Ninnemann O., Janniaux J.-C., Frommer W.B. EMBO J. 13:3464-3471(1994).

[2] Siewe R.M., Weil B., Burkovski A., Eikmanns B.J., Eikmanns M., Kraemer R. J. Biol. Chem. 271:5398-5403(1996).

[3] Saier M.H. Jr. Adv. Microbiol. Physiol. 40:81-136(1998).

29. (Arch_histone)

CBF/NF-Y subunits signatures

Diverse DNA binding proteins are known to bind the CCAAT box, a common cis-acting element found in the promoter and enhancer regions of a large number of genes in eukaryotes. Amongst these proteins is one known as the CCAAT-binding factor (CBF) or NF-Y [1]. CBF is a heteromeric transcription factor that consists of two different components both needed for DNA-binding. The HAP protein complex of yeast binds to the upstream activation site of cytochrome C iso-1 gene (CYC1) as well as other genes involved in mitochondrial electron transport and activates their expression. It also recognizes the sequence CCAAT and is structurally and evolutionary related to CBF. The first subunit of CBF, known as CBF-A or NF-YB in vertebrates, HAP3 in budding yeast and as php3 in fission yeast, is a protein of 116 to 210 amino-acid residues which contains a highly conserved central domain of about 90residues. This domain seems to be involved in DNA-binding; a signature pattern had been developed from its central part. The second subunit of CBF, known as CBF-B or NF-YA in vertebrates, HAP2 in budding yeast and php2 in fission yeast, is a protein of 265 to 350 amino-acid residues which contains a highly conserved region of about 60 residues. This region, called the 'essential core' [2], seems to consist of two subdomains: an N-terminal subunit-association domain and a C-terminal DNA recognition domain. A signature pattern has been developed from a section of the subunit-association domain.

Consensus pattern: C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C-

Consensus pattern: Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E-

[1] Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I., Benoist C., Mathis D. Nucleic Acids Res. 20:1087-1091(1992).

[2] Olesen J.T., Fikes J.D., Guarente L. Mol. Cell. Biol. 11:611-619(1991).

30. Argininosuccinate synthase signatures

Argininosuccinate synthase (EC 6.3.4.5) (AS) is a urea cycle enzyme that catalyzes the penultimate step in arginine biosynthesis: the ATP-dependent ligation of citrulline to aspartate to form argininosuccinate, AMP andpyrophosphate [1,2].In humans, a defect in the AS gene causes citrullinemia, a genetic disease characterized by severe vomiting spells and mental retardation.AS is a homotetrameric enzyme of chains of about 400 amino-acid residues. Anarginine seems to be important for the enzyme's catalytic mechanism. The sequences of AS from various prokaryotes, archaebacteria and eukaryotes show significant similarity. Two signature patterns have been selected for AS. The first is a highly conserved stretch of nine residues located in the N-terminal extremity of these enzymes, the second is derived from a conserved region which contains one of the conserved arginine residues.

Consensus pattern: [AS]-[FY]-S-G-G-[LV]-D-T-[ST]-

Consensus pattern: G-x-T-x-K-G-N-D-x(2)-R-F-

[1] van Vliet F., Crabeel M., Boyen A., Tricot C., Stalon V., Falmagne P., Nakamura Y., Baumberg S., Glansdorff N. Gene 95:99-104(1990).
[2] Morris C.J., Reeve J.N. J. Bacteriol. 170:3125-3130(1988).

31. Armadillo/beta-catenin-like repeats

Approx. 40 amino acid repeat. Tandem repeats form super-helix of helices that is proposed to mediate interaction of beta-catenin with its ligands. CAUTION: This family does not contain all known armadillo repeats.

[1] Huber AH, Nelson WJ, Weis WI, Cell 1997;90:871-882.
[2] Gumbiner BM, Curr Opin Cell Biol 1995;7:634-640.
[3] Cavallo R, Rubenstein D, Peifer M, Curr Opin Genet Dev 1997;7:459-466.
[4] Su LK, Vogelstein B, Kinzler KW, Science 1993;262:1734-1737.

[5] Masiarz FR, Munemitsu S, Polakis P Science 1993;262:1731-1734

[6] Peifer M, Wieschaus E, Cell 1990;63:1167-1176.

32. (Asn Synthase)

Asparagine synthase

This family is always found associated with GATase_2. Members of this family catalyse the conversion of aspartate to asparagine.

33. Asparaginase_2

Asparaginase 12 members

34. (Aspartyl tRNA N)

Aminoacyl-transfer RNA synthetases class-II signatures

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).
[2] Delarue M., Moras D. BioEssays 15:675-687(1993).
[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).
[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).
[5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).
[6] Cusack S. Biochimie 75:1077-1081(1993).
[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).
[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

35. (ArfGap) Putative GTP-ase activating protein for Arf. Putative zinc fingers with GTPase activating proteins (GAPs) towards the small GTPase, Arf. The GAP of ARD1 stimulates GTPase hydrolysis for ARD1 but not ARFs. Number of members: 34

[1]Medline: 96324970. Identification and cloning of centaurin-alpha. A novel phosphatidylinositol 3,4,5-trisphosphate-binding protein from rat brain. Hammonds-Odie LP, Jackson TR, Profit AA, Blader IJ, Turck CW, Prestwich GD, Theibert AB; J Biol Chem 1996;271:18859-18868.
[2]Medline: 97296423. A target of phosphatidylinositol 3,4,5-trisphosphate with a zinc finger motif similar to that of the ADP-ribosylation -factor GTPase-activating protein and two pleckstrin homology domains. Tanaka K, Imajoh-Ohmi S, Sawada T, Shirai R, Hashimoto Y, Iwasaki S, Kaibuchi K, Kanaho Y, Shirai T, Terada Y, Kimura K, Nagata S, Fukui Y; Eur J Biochem 1997;245:512-519.
[3] 98112795. Molecular characterization of the GTPase-activating domain of ADP-ribosylation factor domain protein 1 (ARD1). Vitale N, Moss J, Vaughan M; J Biol Chem 1998;273:2553-2560.

36. Apolipoprotein. Apolipoprotein A1/A4/E family. This family includes: Swiss:P02647 Apolipoprotein A-I. Swiss:P06727 Apolipoprotein A-IV. Swiss:P02649 Apolipoprotein E.

These proteins contain several 22 residue repeats which form a pair of alpha helices. Number of members: 42

[1]Medline: 91289138. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA; Science 1991;252:1817-1822.

37. Amino acid permeases signature

Amino acid permeases are integral membrane proteins involved in the transport of amino acids into the cell. A number of such proteins have been found to be evolutionary related [1,2,3]. These proteins are: - Yeast general amino acid permeases (genes GAP1, AGP2 and AGP3). - Yeast basic amino acid permease (gene ALP1). - Yeast Leu/Val/Ile permease (gene BAP2). - Yeast arginine permease (gene CAN1). - Yeast dicarboxylic amino acid permease (gene DIP5). - Yeast asparagine/glutamine permease (gene AGP1). - Yeast glutamine permease (gene GNP1). - Yeast histidine permease (gene HIP1). - Yeast lysine permease (gene LYP1). - Yeast proline permease (gene PUT4). - Yeast valine and tyrosine permease (gene VAL1/TAT1). - Yeast tryptophan permease (gene TAT2/SCM2). - Yeast choline transport protein (gene HNM1/CTR1). - Yeast GABA permease (gene UGA4). - Yeast hypothetical protein YKL174c. - Fission yeast protein isp5. - Fission yeast hypothetical protein SpAC8A4.11 - Fission yeast hypothetical protein SpAC11D3.08c. - Emericella nidulans proline transport protein (gene prnB). - Trichoderma harzianum amino acid permease INDA1. - Salmonella typhimurium L-asparagine permease (gene ansP). - Escherichia coli aromatic amino acid transport protein (gene aroP). - Escherichia coli D-serine/D-alanine/glycine transporter (gene cycA). - Escherichia coli GABA permease (gene gabP). - Escherichia coli lysine-specific permease (gene lysP). - Escherichia coli phenylalanine-specific permease (gene pheP). - Salmonella typhimurium proline-specific permease (gene proY). - Escherichia coli and Klebsiella pneumoniae hypothetical protein yeeF. - Escherichia coli and Salmonella typhimurium hypothetical protein yifK. - Bacillus subtilis permeases rocC and rocE which probably transports arginine or ornithine. These proteins seem to contain up to 12 transmembrane segments. As a signature for this family of proteins, the best conserved region which is located in the second transmembrane segment has been selected.

Consensus pattern: [STAGC]-G-[PAG]-x(2,3)-[LIVMFYWA](2)-x-[LIVMFYW]-x-[LIVMFWSTAGC](2)-[STAGC]-x(3)-[LIVMFYWT]-x-[LIVMST]-x(3)- [LIVMCTA]-[GA]-E-x(5)-[PSAL]-

[1] Weber E., Chevalier M.R., Jund R. J. Mol. Evol. 27:341-350(1988).

[2] Vandenbol M., Jauniaux J.-C., Grenson M. Gene 83:153-159(1989).

[3] Reizer J., Finley K., Kakuda D., McLeod C.L., Reizer A., Saier M.H. Jr. Protein Sci. 2:20-30(1993).

38. aakinase (1) Glutamate 5-kinase signature

Glutamate 5-kinase (EC 2.7.2.11) (gamma-glutamyl kinase) (GK) is the enzyme that catalyzes the first step in the biosynthesis of proline from glutamate, the ATP-dependent phosphorylation of L-glutamate into L-glutamate 5-phosphate. In eubacteria (gene proB) and yeast [1] (gene PRO1), GK is a monofunctional protein, while in plants and mammals, it is a bifunctional enzyme (P5CS) [2]that consists of two domains: a N-terminal GK domain and a C-terminal gamma-glutamyl phosphate reductase domain (EC 1.2.1.41) (see <PDOC00940>).As a signature pattern, a highly conserved glycine-and alanine-rich region located in the central section of these enzymes has been selected. Yeast hypothetical protein YHR033w is highly similar to GK.

Consensus pattern: [GSTN]-x(2)-G-x-G-[GC]-[IM]-x-[STA]-K-[LIVM]-x-[SA]-[TCA]-x(2)-[GALV]-x(3)-G-

[1] Li W., Brandriss M.C. J. Bacteriol. 174:4148-4156(1992).

[2] Hu C.-A.A., Delauney A.J., Verma D.P.S. Proc. Natl. Acad. Sci. U.S.A. 89:9354-9358(1992).

aakinase (2) Aspartokinase signature

Aspartokinase (EC 2.7.2.4) (AK) [1] catalyzes the phosphorylation of aspartate. The product of this reaction can then be used in the biosynthesis of lysine or in the pathway leading to homoserine, which participates in the biosynthesis of threonine, isoleucine and methionine. In Escherichia coli, there are three different isozymes which differ in their sensitivity to repression and inhibition by Lys, Met and Thr. AK1 (gene thrA) and AK2 (gene metL) are

bifunctional enzymes which both consist of an N- terminal AK domain and a C-terminal homoserine dehydrogenase domain. AK1 is involved in threonine biosynthesis and AK2, in that of methionine. The third isozyme, AK3 (gene lysC), is monofunctional and involved in lysine synthesis. In yeast, there is a single isozyme of AK (gene HOM3). As a signature pattern for AK, a conserved region located in the N-terminal extremity has been selected.

Consensus pattern: [LIVM]-x-K-[FY]-G-G-[ST]-[SC]-[LIVM]-

[1] Rafalski J.A., Falco S.C. J. Biol. Chem. 263:2146-2151(1988).

aakinase (3) Gamma-glutamyl phosphate reductase signature

Gamma-glutamyl phosphate reductase (EC 1.2.1.41) (GPR) is the enzyme that catalyzes the second step in the biosynthesis of proline from glutamate, the NADP-dependent reduction of L-glutamate 5-phosphate into L-glutamate 5-semialdehyde and phosphate. In eubacteria (gene proA) and yeast [1] (gene PRO2), GPR is a monofunctional protein, while in plants and mammals, it is a bifunctional enzyme (P5CS) [2]that consists of two domains: a N-terminal glutamate 5-kinase domain(EC 2.7.2.11) (see <PDOC00701>) and a C-terminal GPR domain. As a signature pattern, a conserved region that contains two histidine residues has been selected. This region is located in the last third of GPR.

Consensus pattern: V-x(5)-A-[LIV]-x-H-I-x(2)-[HY]-[GS]-[ST]-x-H-[ST]-[DE]-x- I-

[1] Pearson B.M., Hernando Y., Payne J., Wolf S.S., Kalogeropoulos A., Schweizer M. Yeast 12:1021-1031(1996).

[2] Hu C.-A.A., Delauney A.J., Verma D.P.S. Proc. Natl. Acad. Sci. U.S.A. 89:9354-9358(1992).

39. (abhydrolase) alpha/beta hydrolase fold. This catalytic domain is found in a very wide range of enzymes.

[1] Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KHG, Goldman A, Protein Eng 1992;5:197-211.

40. (Acid phosphat) Histidine acid phosphatases signatures

Acid phosphatases (EC 3.1.3.2) are a heterogeneous group of proteins that hydrolyze phosphate esters, optimally at low pH. It has been shown [1] that a number of acid phosphatases, from both prokaryotes and eukaryotes, share two regions of sequence similarity, each centered around a conserved histidine residue. These two histidines seem to be involved in the enzymes' catalytic mechanism [2,3]. The first histidine is located in the N-terminal section and forms a phosphohistidine intermediate while the second is located in the C- terminal section and possibly acts as proton donor. Enzymes belonging to this family are called 'histidine acid phosphatases' and are listed below:

- Escherichia coli pH 2.5 acid phosphatase (gene appA).

- Escherichia coli glucose-1-phosphatase (EC 3.1.3.10) (gene agp).
- Yeast constitutive and repressible acid phosphatases (genes PHO3 and PHO5).
- Fission yeast acid phosphatase (gene pho1).
- Aspergillus phytases A and B (EC 3.1.3.8) (gene phyA and phyB).
- Mammalian lysosomal acid phosphatase.
- Mammalian prostatic acid phosphatase.
- Caenorhabditis elegans hypothetical proteins B0361.7, C05C10.1, C05C10.4 and F26C11.1.

Consensus pattern[LIVM]-x(2)-[LIVMA]-x(2)-[LIVM]-x-R-H-[GN]-x-R-x-[PAS] [H is the phosphohistidine residue]

Consensus pattern[LIVMF]-x-[LIVMFAG]-x(2)-[STAGI]-H-D-[STANQ]-x-[LIVM]-x(2)-[LIVMFY]-x(2)-[STA] [H is an active site residue] Sequences known to belong to this class detected by the patternALL, except for rat prostatic acid phosphatase which seems to have Tyr instead of the active site His

[1] van Etten R.L., Davidson R., Stevis P.E., MacArthur H., Moore D.L. J. Biol. Chem. 266:2313-2319(1991).

[2] Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M., van Etten R.L. J. Biol. Chem. 267:22830-22836(1992).

[3] Schneider G., Lindqvist Y., Vihko P. EMBO J. 12:2609-2615(1993).

41. Aconitase family signatures

Aconitase (aconitate hydratase) (EC 4.2.1.3) [1] is the enzyme from the tricarboxylic acid cycle that catalyzes the reversible isomerization of citrate and isocitrate. Cis-aconitate is formed as an intermediary product during the course of the reaction. In eukaryotes two isozymes of aconitase are known to exist: one found in the mitochondrial matrix and the other found in the cytoplasm. Aconitase, in its active form, contains a 4Fe-4S iron-sulfur cluster; three cysteine residues have been shown to be ligands of the 4Fe-4S cluster. It has been shown that the aconitase family also contains the following proteins: - Iron-responsive element binding protein (IRE-BP). IRE-BP is a cytosolic protein that binds to iron-responsive elements (IREs). IREs are stem-loop structures found in the 5'UTR of ferritin, and delta aminolevulinic acid synthase mRNAs, and in the 3'UTR of transferrin receptor mRNA. IRE-BP also express aconitase activity. - 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate isomerase), the enzyme that catalyzes the second step in the biosynthesis of leucine. - Homoaconitase (EC 4.2.1.36) (homoaconitate hydratase), an enzyme that participates in the alpha-aminoadipate pathway of lysine biosynthesis and that converts cis-homoaconitate into homoisocitric acid. - Esherichia coli protein ybhJ

Consensus pattern: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-[LIVMA] [C binds the iron-sulfur center]
Consensus pattern: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA] [The two C's bind the iron-sulfur center]-

[1] Gruer M.J., Artymiuk P.J., Guest J.R. Trends Biochem. Sci. 22:3-6(1997).

42. Actins signatures

Actins [1 to 4] are highly conserved contractile proteins that are present in all eukaryotic cells. In vertebrates there are three groups of actin isoforms: alpha, beta and gamma. The alpha actins are found in muscle tissues and are a major constituent of the contractile apparatus. The beta and gamma actins co-exists in most cell types as components of the cytoskeleton and as mediators of internal cell motility. In plants [5] there are many isoforms

which are probably involved in a variety of functions such as cytoplasmic streaming, cell shape determination, tip growth, graviperception, cell wall deposition, etc. Actin exists either in a monomeric form (G-actin) or in a polymerized form (F-actin). Each actin monomer can bind a molecule of ATP; when polymerization occurs, the ATP is hydrolyzed. Actin is a protein of from 374 to 379 amino acid residues. The structure of actin has been highly conserved in the course of evolution. Recently some divergent actin-like proteins have been identified in several species. These proteins are: - Centractin (actin-RPV) from mammals, fungi (yeast ACT5, Neurospora crassa ro-4) and Pneumocystis carinii (actin-II). Centractin seems to be a component of a multi-subunit centrosomal complex involved in microtubule based vesicle motility. This subfamily is also known as ARP1. - ARP2 subfamily which includes chicken ACTL, yeast ACT2, Drosophila 14D, C.elegans actC. - ARP3 subfamily which includes actin 2 from mammals, Drosophila 66B, yeast ACT4 and fission yeast act2. - ARP4 subfamily which includes yeast ACT3 and Drosophila 13E. Three signature patterns have been developed. The first two are specific to actins and span positions 54 to 64 and 357 to 365. The last signature picks up both actins and the actin-like proteins and corresponds to positions 106 to 118 in actins.

Consensus pattern: [FY]-[LIV]-G-[DE]-E-A-Q-x-[RKQ](2)-G-

Consensus pattern: W-[IV]-[STA]-[RK]-x-[DE]-Y-[DNE]-[DE]-

Consensus pattern: [LM]-[LIVM]-T-E-[GAPQ]-x-[LIVMFYWHQ]-N-[PSTAQ]-x(2)-N-[KR]-

[1] Sheterline P., Clayton J., Sparrow J.C. (In) Actins, 3rd Edition, Academic Press Ltd, London, (1996).

[2] Pollard T.D., Cooper J.A. Annu. Rev. Biochem. 55:987-1036(1986).

[3] Pollard T.D. Curr. Opin. Cell Biol. 1:33-40(1990).

[4] Rubenstein P.A. BioEssays 12:309-315(1990).

[5] Meagher R.B., McLean B.G. Cell Motil. Cytoskeleton 16:164-166(1990).

43. Adenylate kinase signature

Adenylate kinase (EC 2.7.4.3) (AK) [1] is a small monomeric enzyme that catalyzes the reversible transfer of MgATP to AMP (MgATP + AMP = MgADP + ADP).In mammals there are three different isozymes: - AK1 (or myokinase), which is cytosolic. - AK2, which is

located in the outer compartment of mitochondria. - AK3 (or GTP:AMP phosphotransferase), which is located in the mitochondrial matrix and which uses MgGTP instead of MgATP.The sequence of AK has also been obtained from different bacterial species and from plants and fungi. Two other enzymes have been found to be evolutionary related to AK. These are: - Yeast uridylate kinase (EC 2.7.4.-) (UK) (gene URA6) [2] which catalyzes the transfer of a phosphate group from ATP to UMP to form UDP and ADP. - Slime mold UMP-CMP kinase (EC 2.7.4.14) [3] which catalyzes the transfer of a phosphate group from ATP to either CMP or UMP to form CDP or UDP and ADP. Several regions of AK family enzymes are well conserved, including the ATP-binding domains. The most conserved of all regions have been selected as a signature for this type of enzyme. This region includes an aspartic acid residue that is part of the catalytic cleft of the enzyme and that is involved in a salt bridge. It also includes an arginine residue whose modification leads to inactivation of the enzyme

Consensus pattern: [LIVMFYW](3)-D-G-[FYI]-P-R-x(3)-[NQ]-

[1] Schulz G.E. Cold Spring Harbor Symp. Quant. Biol. 52:429-439(1987).
[2] Liljelund P., Sanni A., Friesen J.D., Lacroute F. Biochem. Biophys. Res. Commun. 165:464-473(1989).
[3] Wiesmueller L., Noegel A.A., Barzu O., Gerisch G., Schleicher M. J. Biol. Chem. 265:6339-6345(1990).
[4] Kath T.H., Schmid R., Schaefer G. Arch. Biochem. Biophys. 307:405-410(1993).

44. (adh_short) Short-chain dehydrogenases/reductases family signature. The short-chain dehydrogenases/reductases family (SDR) [1] is a very large family of enzymes, most of which are known to be NAD- or NADP-dependent oxidoreductases. As the first member of this family to be characterized was Drosophila alcohol dehydrogenase, this family used to be called [2,3,4]'insect-type', or 'short-chain' alcohol dehydrogenases. Most member of this family are proteins of about 250 to 300 amino acid residues. The proteins currently known to belong to this family are listed below. - Alcohol dehydrogenase (EC 1.1.1.1) from insects such as Drosophila. - Acetoin dehydrogenase (EC 1.1.1.5) from Klebsiella terrigena (gene budC). - D-beta-hydroxybutyrate dehydrogenase (BDH) (EC 1.1.1.30) from mammals. - Acetoacetyl-CoA reductase (EC 1.1.1.36) from various bacterial species (gene phbB or phaB). - Glucose 1-dehydrogenase (EC 1.1.1.47) from Bacillus. - 3-beta-hydroxysteroid

dehydrogenase (EC 1.1.1.51) from Comomonas testosteroni. - 20-beta-hydroxysteroid dehydrogenase (EC 1.1.1.53) from Streptomyces hydrogenans. - Ribitol dehydrogenase (EC 1.1.1.56) (RDH) from Klebsiella aerogenes. - Estradiol 17-beta-dehydrogenase (EC 1.1.1.62) from human. - Gluconate 5-dehydrogenase (EC 1.1.1.69) from Gluconobacter oxydans (gene gno). - 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) from Escherichia coli (gene fabG) and from plants. - Retinol dehydrogenase (EC 1.1.1.105) from mammals. - 2-deoxy-d-gluconate 3-dehydrogenase (EC 1.1.1.125) from Escherichia coli and Erwinia chrysanthemi (gene kduD). - Sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.1.140) from Escherichia coli (gene gutD) and from Klebsiella pneumoniae (gene sorD). - 15-hydroxyprostaglandin dehydrogenase (NAD+) (EC 1.1.1.141) from human. - Corticosteroid 11-beta-dehydrogenase (EC 1.1.1.146) (11-DH) from mammals. - 7-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) from Escherichia coli (gene hdhA), Eubacterium strain VPI 12708 (gene baiA) and from Clostridium sordellii. - NADPH-dependent carbonyl reductase (EC 1.1.1.184) from mammals. - Tropinone reductase-I (EC 1.1.1.206) and -II (EC 1.1.1.236) from plants. - N-acylmannosamine 1-dehydrogenase (EC 1.1.1.233) from Flavobacterium strain 141-8. - D-arabinitol 2-dehydrogenase (ribulose forming) (EC 1.1.1.250) from fungi. - Tetrahydroxynaphthalene reductase (EC 1.1.1.252) from Magnaporthe grisea. - Pteridine reductase 1 (EC 1.1.1.253) (gene PTR1) from Leishmania. - 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-) from Pseudomonas paucimobilis. - Cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase (EC 1.3.1. -) from Acinetobacter calcoaceticus (gene benD) and Pseudomonas putida (gene xylL). - Biphenyl-2,3-dihydro-2,3-diol dehydrogenase (EC 1.3.1.-) (gene bphB) from various Pseudomonaceae. - Cis-toluene dihydrodiol dehydrogenase (EC 1.3.1.-) from Pseudomonas putida (gene todD). - Cis-benzene glycol dehydrogenase (EC 1.3.1.19) from Pseudomonas putida (gene bnzE). - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28) from Escherichia coli (gene entA) and Bacillus subtilis (gene dhbA). - Dihydropteridine reductase (EC 1.6.99.7) (HDHPR) from mammals. - Lignin degradation enzyme ligD from Pseudomonas paucimobilis. - Agropine synthesis reductase from Agrobacterium plasmids (gene mas1). - Versicolorin reductase from Aspergillus parasiticus (gene VER1). - Putative keto-acyl reductases from Streptomyces polyketide biosynthesis operons. - A trifunctional hydratase-dehydrogenase-epimerase from the peroxisomal beta-oxidation system of Candida tropicalis. This protein contains two tandemly repeated 'short-chain dehydrogenase-type' domain in its N-terminal extremity. - Nodulation protein nodG from species of Azospirillum and Rhizobium which is probably involved in the modification of the nodulation Nod factor fatty

acyl chain. - Nitrogen fixation protein fixR from Bradyrhizobium japonicum. - Bacillus subtilis protein dltE which is involved in the biosynthesis of D- alanyl-lipoteichoic acid. - Human follicular variant translocation protein 1 (FVT1). - Mouse adipocyte protein p27. - Mouse protein Ke 6. - Maize sex determination protein TASSELSEED 2. - Sarcophaga peregrina 25 Kd development specific protein. - Drosophila fat body protein P6. - A Listeria monocytogenes hypothetical protein encoded in the internalins gene region. - Escherichia coli hypothetical protein yciK. - Escherichia coli hypothetical protein ydfG. - Escherichia coli hypothetical protein yjgI. - Escherichia coli hypothetical protein yjgU. - Escherichia coli hypothetical protein yohF. - Bacillus subtilis hypothetical protein yoxD. - Bacillus subtilis hypothetical protein ywfD. - Bacillus subtilis hypothetical protein ywfH. - Yeast hypothetical protein YIL124w. - Yeast hypothetical protein YIR035c. - Yeast hypothetical protein YIR036c. - Yeast hypothetical protein YKL055c. - Fission yeast hypothetical protein SpAC23D3.11. One of the best conserved regions which includes two perfectly conserved residues, a tyrosine and a lysine has been selected as a signature pattern for this family of proteins. The tyrosine residue participates in the catalytic mechanism.

Consensus pattern: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-{PC}-[SAGFYR]-[LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)- [LIVMFYWGAPTHQ]-[GSACQRHM] [Y is an active site residue] -

[1] Joernvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D. Biochemistry 34:6003-6013(1995).
[2] Villarroya A., Juan E., Egestad B., Joernvall H. Eur. J. Biochem. 180:191-197(1989).
[3] Persson B., Krook M., Joernvall H. Eur. J. Biochem. 200:537-543(1991).
[4] Neidle E.L., Hartnett C., Ornston N.L., Bairoch A., Rekik M., Harayama S. Eur. J. Biochem. 204:113-120(1992).

45. (adh_short_C2) Short-chain dehydrogenases/reductases family signature

The short-chain dehydrogenases/reductases family (SDR) [1] is a very large family of enzymes, most of which are known to be NAD- or NADP-dependent oxidoreductases. As the first member of this family to be characterized was Drosophila alcohol dehydrogenase, this family used to be called [2,3,4]'insect-type', or 'short-chain' alcohol dehydrogenases. Most member of this family are proteins of about 250 to 300 amino acid residues. The proteins

currently known to belong to this family are listed below. - Alcohol dehydrogenase (EC 1.1.1.1) from insects such as Drosophila. - Acetoin dehydrogenase (EC 1.1.1.5) from Klebsiella terrigena (gene budC). - D-beta-hydroxybutyrate dehydrogenase (BDH) (EC 1.1.1.30) from mammals. - Acetoacetyl-CoA reductase (EC 1.1.1.36) from various bacterial species (gene phbB or phaB). - Glucose 1-dehydrogenase (EC 1.1.1.47) from Bacillus. - 3-beta-hydroxysteroid dehydrogenase (EC 1.1.1.51) from Comomonas testosteroni. - 20-beta-hydroxysteroid dehydrogenase (EC 1.1.1.53) from Streptomyces hydrogenans. - Ribitol dehydrogenase (EC 1.1.1.56) (RDH) from Klebsiella aerogenes. - Estradiol 17-beta-dehydrogenase (EC 1.1.1.62) from human. - Gluconate 5-dehydrogenase (EC 1.1.1.69) from Gluconobacter oxydans (gene gno). - 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) from Escherichia coli (gene fabG) and from plants. - Retinol dehydrogenase (EC 1.1.1.105) from mammals. - 2-deoxy-d-gluconate 3-dehydrogenase (EC 1.1.1.125) from Escherichia coli and Erwinia chrysanthemi (gene kduD). - Sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.1.140) from Escherichia coli (gene gutD) and from Klebsiella pneumoniae (gene sorD). - 15-hydroxyprostaglandin dehydrogenase (NAD+) (EC 1.1.1.141) from human. - Corticosteroid 11-beta-dehydrogenase (EC 1.1.1.146) (11-DH) from mammals. - 7-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) from Escherichia coli (gene hdhA), Eubacterium strain VPI 12708 (gene baiA) and from Clostridium sordellii. - NADPH-dependent carbonyl reductase (EC 1.1.1.184) from mammals. - Tropinone reductase-I (EC 1.1.1.206) and -II (EC 1.1.1.236) from plants. - N-acylmannosamine 1-dehydrogenase (EC 1.1.1.233) from Flavobacterium strain 141-8. - D-arabinitol 2-dehydrogenase (ribulose forming) (EC 1.1.1.250) from fungi. - Tetrahydroxynaphthalene reductase (EC 1.1.1.252) from Magnaporthe grisea. - Pteridine reductase 1 (EC 1.1.1.253) (gene PTR1) from Leishmania. - 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-) from Pseudomonas paucimobilis. - Cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase (EC 1.3.1. -) from Acinetobacter calcoaceticus (gene benD) and Pseudomonas putida (gene xylL). - Biphenyl-2,3-dihydro-2,3-diol dehydrogenase (EC 1.3.1.-) (gene bphB) from various Pseudomonaceae. - Cis-toluene dihydrodiol dehydrogenase (EC 1.3.1.-) from Pseudomonas putida (gene todD). - Cis-benzene glycol dehydrogenase (EC 1.3.1.19) from Pseudomonas putida (gene bnzE). - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28) from Escherichia coli (gene entA) and Bacillus subtilis (gene dhbA). - Dihydropteridine reductase (EC 1.6.99.7) (HDHPR) from mammals. - Lignin degradation enzyme ligD from Pseudomonas paucimobilis. - Agropine synthesis reductase from Agrobacterium plasmids (gene mas1). - Versicolorin reductase from Aspergillus

parasiticus (gene VER1). - Putative keto-acyl reductases from Streptomyces polyketide biosynthesis operons. - A trifunctional hydratase-dehydrogenase-epimerase from the peroxisomal beta-oxidation system of Candida tropicalis. This protein contains two tandemly repeated 'short-chain dehydrogenase-type' domain in its N-terminal extremity. - Nodulation protein nodG from species of Azospirillum and Rhizobium which is probably involved in the modification of the nodulation Nod factor fatty acyl chain. - Nitrogen fixation protein fixR from Bradyrhizobium japonicum. - Bacillus subtilis protein dltE which is involved in the biosynthesis of D- alanyl-lipoteichoic acid. - Human follicular variant translocation protein 1 (FVT1). - Mouse adipocyte protein p27. - Mouse protein Ke 6. - Maize sex determination protein TASSELSEED 2. - Sarcophaga peregrina 25 Kd development specific protein. - Drosophila fat body protein P6. - A Listeria monocytogenes hypothetical protein encoded in the internalins gene region. - Escherichia coli hypothetical protein yciK. - Escherichia coli hypothetical protein ydfG. - Escherichia coli hypothetical protein yjgI. - Escherichia coli hypothetical protein yjgU. - Escherichia coli hypothetical protein yohF. - Bacillus subtilis hypothetical protein yoxD. - Bacillus subtilis hypothetical protein ywfD. - Bacillus subtilis hypothetical protein ywfH. - Yeast hypothetical protein YIL124w. - Yeast hypothetical protein YIR035c. - Yeast hypothetical protein YIR036c. - Yeast hypothetical protein YKL055c. - Fission yeast hypothetical protein SpAC23D3.11. One of the best conserved regions which includes two perfectly conserved residues, a tyrosine and a lysine has been used as a signature pattern for this family of proteins. The tyrosine residue participates in the catalytic mechanism.

Consensus pattern: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-{PC}-[SAGFYR]-[LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)- [LIVMFYWGAPTHQ]-[GSACQRHM] [Y is an active site residue]

[1] Joernvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D. Biochemistry 34:6003-6013(1995).

[2] Villarroya A., Juan E., Egestad B., Joernvall H. Eur. J. Biochem. 180:191-197(1989).

[3] Persson B., Krook M., Joernvall H. Eur. J. Biochem. 200:537-543(1991).

[4] Neidle E.L., Hartnett C., Ornston N.L., Bairoch A., Rekik M., Harayama S. Eur. J. Biochem. 204:113-120(1992).

46. (adh_zinc) Zinc-containing alcohol dehydrogenases signatures

Alcohol dehydrogenase (EC 1.1.1.1) (ADH) catalyzes the reversible oxidation of ethanol to acetaldehyde with the concomitant reduction of NAD [1]. Currently three, structurally and catalytically, different types of alcohol dehydrogenases are known: - Zinc-containing 'long-chain' alcohol dehydrogenases. - Insect-type, or 'short-chain' alcohol dehydrogenases. - Iron-containing alcohol dehydrogenases.Zinc-containing ADH's [2,3] are dimeric or tetrameric enzymes that bind two atoms of zinc per subunit. One of the zinc atom is essential for catalytic activity while the other is not. Both zinc atoms are coordinated by either cysteine or histidine residues; the catalytic zinc is coordinated by two cysteines and one histidine. Zinc-containing ADH's are found in bacteria, mammals, plants, and in fungi. In most species there are more than one isozyme (for example, human have at least six isozymes, yeast have three, etc.). A number of other zinc-dependent dehydrogenases are closely related to zinc ADH [4], these are: - Xylitol dehydrogenase (EC 1.1.1.9) (D-xylulose reductase). - Sorbitol dehydrogenase (EC 1.1.1.14). - Aryl-alcohol dehydrogenase (EC 1.1.1.90) (benzyl alcohol dehydrogenase). - Threonine 3-dehydrogenase (EC 1.1.1.103). - Cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) (CAD) [5]. CAD is a plant enzyme involved in the biosynthesis of lignin. - Galactitol-1-phosphate dehydrogenase (EC 1.1.1.251). - Pseudomonas putida 5-exo-alcohol dehydrogenase (EC 1.1.1.-) [6]. - Escherichia coli starvation sensing protein rspB. - Escherichia coli hypothetical protein yjgB. - Escherichia coli hypothetical protein yjgV. - Escherichia coli hypothetical protein yjjN. - Yeast hypothetical protein YAL060w (FUN49). - Yeast hypothetical protein YAL061w (FUN50). - Yeast hypothetical protein YCR105w. The pattern that has been developed to detect this class of enzymes is based on a conserved region that includes a histidine residue which is the second ligand of the catalytic zinc atom. This family also includes NADP-dependent quinone oxidoreductase (EC 1.6.5.5),an enzyme found in bacteria (gene qor), in yeast and in mammals where, in some species such as rodents, it has been recruited as an eye lens protein and is known as zeta-crystallin [7]. The sequence of quinone oxidoreductase is distantly related to that other zinc-containing alcohol dehydrogenases and it lacks the zinc-ligand residues. The torpedo fish and mammlian synaptic vesicle membrane protein vat-1 is related to qor. A specific pattern has been developed for this subfamily.

Consensus pattern: G-H-E-x(2)-G-x(5)-[GA]-x(2)-[IVSAC] [H is a zinc ligand]

Consensus pattern: [GSD]-[DEQH]-x(2)-L-x(3)-[SA](2)-G-G-x-G-x(4)-Q-x(2)-[KR]-

[1] Branden C.-I., Joernvall H., Eklund H., Furugren B. (In) The Enzymes (3rd edition) 11:104-190(1975).

[2] Joernvall H., Persson B., Jeffery J. Eur. J. Biochem. 167:195-201(1987).

[3] Sun H.-W., Plapp B.V. J. Mol. Evol. 34:522-535(1992).

[4] Persson B., Hallborn J., Walfridsson M., Hahn-Haegerdal B., Keraenen S., Penttilae M., Joernvall H. FEBS Lett. 324:9-14(1993).

[5] Knight M.E., Halpin C., Schuch W. Plant Mol. Biol. 19:793-801(1992).

[6] Koga H., Aramaki H., Yamaguchi E., Takeuchi K., Horiuchi T., Gunsalus I.C. J. Bacteriol. 166:1089-1095(1986).

[7] Joernvall H., Persson B., Du Bois G., Lavers G.C., Chen J.H., Gonzalez P., Rao P.V., Zigler J.S. Jr. FEBS Lett. 322:240-244(1993).

47. (aldedh) Aldehyde dehydrogenases active sites

Aldehyde dehydrogenases (EC 1.2.1.3 and EC 1.2.1.5) are enzymes which oxidize a wide variety of aliphatic and aromatic aldehydes. In mammals at least four different forms of the enzyme are known [1]: class-1 (or Ald C) a tetrameric cytosolic enzyme, class-2 (or Ald M) a tetrameric mitochondrial enzyme, class-3 (or Ald D) a dimeric cytosolic enzyme, and class IV a microsomal enzyme. Aldehyde dehydrogenases have also been sequenced from fungal and bacterial species. A number of enzymes are known to be evolutionary related to aldehyde dehydrogenases; these enzymes are listed below. - Plants and bacterial betaine-aldehyde dehydrogenase (EC 1.2.1.8) [2], an enzyme that catalyzes the last step in the biosynthesis of betaine. - Plants and bacterial NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.9). - Escherichia coli succinate-semialdehyde dehydrogenase (NADP+) (EC 1.2.1.16) (gene gabD) [3], which reduces succinate semialdehyde into succinate. - Escherichia coli lactaldehyde dehydrogenase (EC 1.2.1.22) (gene ald) [4]. - Mammalian succinate semialdehyde dehydrogenase (NAD+) (EC 1.2.1.24). - Escherichia coli phenylacetaldehyde dehydrogenase (EC 1.2.1.39). - Escherichia coli 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (gene hpcC). - Pseudomonas putida 2-hydroxymuconic semialdehyde dehydrogenase [5] (genes dmpC and xylG), an enzyme in the meta-cleavage pathway for the degradation of phenols, cresols and catechol. - Bacterial and mammalian methylmalonate-semialdehyde dehydrogenase (MMSDH) (EC 1.2.1.27) [6], an enzyme involved in the distal pathway of valine catabolism. - Yeast delta-1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12) [7] (gene PUT2), which converts proline to

glutamate. - Bacterial multifunctional putA protein, which contains a delta-1-pyrroline- 5-carboxylate dehydrogenase domain. - 26G, a garden pea protein of unknown function which is induced by dehydration of shoots [8]. - Mammalian formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) [9]. This is a cytosolic enzyme responsible for the NADP-dependent decarboxylative reduction of 10-formyltetrahydrofolate into tetrahydrofolate. It is an protein of about 900 amino acids which consist of three domains; the C- terminal domain (480 residues) is structurally and functionally related to aldehyde dehydrogenases. - Yeast hypothetical protein YBR006w. - Yeast hypothetical protein YER073w. - Yeast hypothetical protein YHR039c. - Caenorhabditis elegans hypothetical protein F01F1.6.A glutamic acid and a cysteine residue have been implicated in the catalytic activity of mammalian aldehyde dehydrogenase. These residues are conserved in all the enzymes of this family. Two patterns have been derived for this family, one for each of the active site residues.

Consensus pattern: [LIVMFGA]-E-[LIMSTAC]-[GS]-G-[KNLM]-[SADN]-[TAPFV] [E is the active site residue]-

Consensus pattern: [FYLVA]-x(3)-G-[QE]-x-C-[LIVMGSTANC]-[AGCN]-x-[GSTADNEKR] [C is the active site residue

[1] Hempel J., Harper K., Lindahl R. Biochemistry 28:1160-1167(1989).
[2] Weretilnyk E.A., Hanson A.D. Proc. Natl. Acad. Sci. U.S.A. 87:2745-2749(1990).
[3] Niegemann E., Schulz A., Bartsch K. Arch. Microbiol. 160:454-460(1993).
[4] Hidalgo E., Chen Y.-M., Lin E.C.C., Aguilar J. J. Bacteriol. 173:6118-6123(1991).
[5] Nordlund I., Shingler V. Biochim. Biophys. Acta 1049:227-230(1990).
[6] Steele M.I., Lorenz D., Hatter K., Park A., Sokatch J.R. J. Biol. Chem. 267:13585-13592(1992).
[7] Krzywicki K.A., Brandriss M.C. Mol. Cell. Biol. 4:2837-2842(1984).
[8] Guerrero F.D., Jones J.T., Mullet J.E. Plant Mol. Biol. 15:11-26(1990).
[9] Cook R.J., Lloyd R.S., Wagner C. J. Biol. Chem. 266:4965-4973(1991).

48. Aldo/keto reductase family signatures

The aldo-keto reductase family [1,2] groups together a number of structurally and functionally related NADPH-dependent oxidoreductases as well as some other proteins. The proteins known to belong to this family are: - Aldehyde reductase (EC 1.1.1.2). - Aldose

reductase (EC 1.1.1.21). - 3-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.50), which terminates androgen action by converting 5-alpha-dihydrotestosterone to 3-alpha-androstanediol. - Prostaglandin F synthase (EC 1.1.1.188) which catalyzes the reduction of prostaglandins H2 and D2 to F2-alpha. - D-sorbitol-6-phosphate dehydrogenase (EC 1.1.1.200) from apple. - Morphine 6-dehydrogenase (EC 1.1.1.218) from Pseudomonas putida plasmid pMDH7.2 (gene morA). - Chlordecone reductase (EC 1.1.1.225) which reduces the pesticide chlordecone (kepone) to the corresponding alcohol. - 2,5-diketo-D-gluconic acid reductase (EC 1.1.1.-) which catalyzes the reduction of 2,5-diketogluconic acid to 2-keto-L-gulonic acid, a key intermediate in the production of ascorbic acid. - NAD(P)H-dependent xylose reductase (EC 1.1.1.-) from the yeast Pichia stipitis. This enzyme reduces xylose into xylit. - Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase (EC 1.3.1.20). - 3-oxo-5-beta-steroid 4-dehydrogenase (EC 1.3.99.6) which catalyzes the reduction of delta(4)-3-oxosteroids. - A soybean reductase, which co-acts with chalcone synthase in the formation of 4,2',4'-trihydroxychalcone. - Frog eye lens rho crystallin. - Yeast GCY protein, whose function is not known. - Leishmania major P110/11E protein. P110/11E is a developmentally regulated protein whose abundance is markedly elevated in promastigotes compared with amastigotes. Its exact function is not yet known. - Escherichia coli hypothetical protein yafB. - Escherichia coli hypothetical protein yghE. - Yeast hypothetical protein YBR149w. - Yeast hypothetical protein YHR104w. - Yeast hypothetical protein YJR096w.These proteins have all about 300 amino acid residues. Three consensus patterns have been developed that are specific to this family of proteins. The first one is located in the N-terminal section of these proteins. The second pattern is located in the central section. The third pattern, located in the C-terminal, is centered on a lysine residue whose chemical modification, in aldose and aldehydereductases, affect the catalytic efficiency.

Consensus pattern: G-[FY]-R-[HSAL]-[LIVMF]-D-[STAGC]-[AS]-x(5)-E-x(2)-[LIVM]- G -
Consensus pattern: [LIVMFY]-x(9)-[KREQ]-x-[LIVM]-G-[LIVM]-[SC]-N-[FY]-
Consensus pattern: [LIVM]-[PAIV]-[KR]-[ST]-x(4)-R-x(2)-[GSTAEQK]-[NSL]-x(2)-[LIVMFA] [K is a putative active site residue]-

[1] Bohren K.M., Bullock B., Wermuth B., Gabbay K.H. J. Biol. Chem. 264:9547-9551(1989).

[2] Bruce N.C., Willey D.L., Coulson A.F.W., Jeffery J. Biochem. J. 299:805-811(1994).

49. Alpha amylase. This family is classified as family 13 of the glycosyl hydrolases. The structure is an 8 stranded alpha/beta barrel, interrupted by a ~70 a.a. calcium-binding domain protruding between beta strand 3 and alpha helix 3, and a carboxyl-terminal Greek key beta-barrel domain.

[1] Larson SB, Greenwood A, Cascio D, Day J, McPherson A, J Mol Biol 1994;235:1560-1584.

50. Aminotransferases class-I pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-I, currently consists of the following enzymes: - Aspartate aminotransferase (AAT) (EC 2.6.1.1). AAT catalyzes the reversible transfer of the amino group from L-aspartate to 2-oxoglutarate to form oxaloacetate and L-glutamate. In eukaryotes, there are two AAT isozymes: one is located in the mitochondrial matrix, the second is cytoplasmic. In prokaryotes, only one form of AAT is found (gene aspC). - Tyrosine aminotransferase (EC 2.6.1.5) which catalyzes the first step in tyrosine catabolism by reversibly transferring its amino group to 2- oxoglutarate to form 4-hydroxyphenylpyruvate and L-glutamate. - Aromatic aminotransferase (EC 2.6.1.57) involved in the synthesis of Phe, Tyr, Asp and Leu (gene tyrB). - 1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) (ACC synthase) from plants. ACC synthase catalyzes the first step in ethylene biosynthesis. - Pseudomonas denitrificans cobC, which is involved in cobalamin biosynthesis. - Yeast hypothetical protein YJL060w.The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [GS]-[LIVMFYTAC]-[GSTA]-K-x(2)-[GSALVN]-[LIVMFA]-x-[GNAR]- x-R-[LIVMA]-[GA] [K is the pyridoxal-P attachment site]

[1] Bairoch A. Unpublished observations (1992).

[2] Sung M.H., Tanizawa K., Tanaka H., Kuramitsu S., Kagamiyama H., Hirotsu K., Okamoto A., Higuchi T., Soda K. J. Biol. Chem. 266:2567-2572(1991).

51. Aminotransferases class-II pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1] into subfamilies. One of these, called class-II, currently consists of the following enzymes: - Glycine acetyltransferase (EC 2.3.1.29), which catalyzes the addition of acetyl-CoA to glycine to form 2-amino-3-oxobutanoate (gene kbl). - 5-aminolevulinic acid synthase (EC 2.3.1.37) (delta-ALA synthase), which catalyzes the first step in heme biosynthesis via the Shemin (or C4) pathway, i.e. the addition of succinyl-CoA to glycine to form 5-aminolevulinate. - 8-amino-7-oxononanoate synthase (EC 2.3.1.47) (7-KAP synthetase), a bacterial enzyme (gene bioF) which catalyzes an intermediate step in the biosynthesis of biotin: the addition of 6-carboxy-hexanoyl-CoA to alanine to form 8-amino-7-oxononanoate. - Histidinol-phosphate aminotransferase (EC 2.6.1.9), which catalyzes the eighth step in histidine biosynthetic pathway: the transfer of an amino group from 3-(imidazol-4-yl)-2-oxopropyl phosphate to glutamic acid to form histidinol phosphate and 2-oxoglutarate. - Serine palmitoyltransferase (EC 2.3.1.50) from yeast (genes LCB1 and LCB2), which catalyzes the condensation of palmitoyl-CoA and serine to form 3- ketosphinganine.The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern

Consensus pattern: T-[LIVMFYW]-[STAG]-K-[SAG]-[LIVMFYWR]-[SAG]-x(2)-[SAG] [K is the pyridoxal-P attachment site]-

[1] Bairoch A. Unpublished observations (1991).

52. Aminotransferases class-III pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped

[1,2] into subfamilies. One of these, called class-III, currently consists of the following enzymes: - Acetylornithine aminotransferase (EC 2.6.1.11) which catalyzes the transfer of an amino group from acetylornithine to alpha-ketoglutarate, yielding N-acetyl-glutamic-5-semi-aldehyde and glutamic acid. - Ornithine aminotransferase (EC 2.6.1.13), which catalyzes the transfer of an amino group from ornithine to alpha-ketoglutarate, yielding glutamic-5- semi-aldehyde and glutamic acid. - Omega-amino acid--pyruvate aminotransferase (EC 2.6.1.18), which catalyzes transamination between a variety of omega-amino acids, mono- and diamines, and pyruvate. It plays a pivotal role in omega amino acids metabolism. - 4-aminobutyrate aminotransferase (EC 2.6.1.19) (GABA transaminase), which catalyzes the transfer of an amino group from GABA to alpha-ketoglutarate, yielding succinate semialdehyde and glutamic acid. - DAPA aminotransferase (EC 2.6.1.62), a bacterial enzyme (gene bioA) which catalyzes an intermediate step in the biosynthesis of biotin, the transamination of 7-keto-8-aminopelargonic acid (7-KAP) to form 7,8- diaminopelargonic acid (DAPA). - 2,2-dialkylglycine decarboxylase (EC 4.1.1.64), a Pseudomonas cepacia enzyme (gene dgdA) that catalyzes the decarboxylating amino transfer of 2,2-dialkylglycine and pyruvate to dialkyl ketone, alanine and carbon dioxide. - Glutamate-1-semialdehyde aminotransferase (EC 5.4.3.8) (GSA). GSA is the enzyme involved in the second step of porphyrin biosynthesis, via the C5 pathway. It transfers the amino group on carbon 2 of glutamate-1- semialdehyde to the neighbouring carbon, to give delta-aminolevulinic acid. - Bacillus subtilis aminotransferase yhxA. - Bacillus subtilis aminotransferase yodT. - Haemophilus influenzae aminotransferase HI0949. - Caenorhabditis elegans aminotransferase T01B11.2.The sequence around the pyridoxal-phosphate attachment site of this class ofenzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [LIVMFYWC](2)-x-D-E-[IVA]-x(2)-G-[LIVMFAGC]-x(0,1)-[RSACLI]-x-[GSAD]-x(12,16)-D-[LIVMFC]-[LIVMFYSTA]-x(2)- [GSA]-K-x(3)-[GSTADNV]-[GSAC] [K is the pyridoxal-P attachment site]-

[1] Bairoch A. Unpublished observations (1992).[2] Yonaha K., Nishie M., Aibara S. J. Biol. Chem. 267:12506-12510(1992).

53. Ank repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure.

[1] A, Holak TA, FEBS Lett 1997;401:127-132.
[2] Lux SE, John KM, Bennett V, Nature 1990;345:736-739.

54. Aminotransferases class-IV signature

Aminotransferases share certain mechanistic features with other pyridoxal-phosphate dependent enzymes, such as the covalent binding of the pyridoxal-phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-IV, currently consists of the following enzymes:

- Branched-chain amino-acid aminotransferase (EC 2.6.1.42) (transaminase B), a bacterial (gene ilvE) and eukaryotic enzyme which catalyzes the reversible transfer of an amino group from 4-methyl-2-oxopentanoate to glutamate, to form leucine and 2-oxoglutarate.
- D-alanine aminotransferase (EC 2.6.1.21). A bacterial enzyme which catalyzes the transfer of the amino group from D-alanine (and other D-amino acids) to 2-oxoglutarate, to form pyruvate and D-aspartate.
- 4-amino-4-deoxychorismate (ADC) lyase (gene pabC). A bacterial enzyme that converts ADC into 4-aminobenzoate (PABA) and pyruvate.

The above enzymes are proteins of about 270 to 415 amino-acid residues that share a few regions of sequence similarity. Surprisingly, the best-conserved region does not include the lysine residue to which the pyridoxal-phosphategroup is known to be attached, in ilvE. The region that has been selected as a signature pattern is located some 40 residues at the C-terminus side of the PIP-lysine

Consensus pattern: E-x-[STAGCI]-x(2)-N-[LIVMFAC]-[FY]-x(6,12)-[LIVMF]-x-T- x(6,8)-[LIVM]-x-[GS]-[LIVM]-x-[KR]-

[1] Green J.M., Merkel W.K., Nichols B.P. J. Bacteriol. 174:5317-5323(1992).
[2] Bairoch A. Unpublished observations (1992).

55. Aminotransferases class-V pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-V, currently consists of the following enzymes: - Phosphoserine aminotransferase (EC 2.6.1.52), an enzyme which catalyzes the reversible interconversion of phosphoserine and 2-oxoglutarate to 3-phosphonooxypyruvate and glutamate. It is required both in the major phosphorylated pathway of serine biosynthesis and in pyridoxine biosynthesis. The bacterial enzyme (gene serC) is highly similar to a rabbit endometrial progesterone-induced protein (EPIP), which is probably a phosphoserine aminotransferase [3]. - Serine--glyoxylate aminotransferase (EC 2.6.1.45) (SGAT) (gene sgaA) from Methylobacterium extorquens. - Serine--pyruvate aminotransferase (EC 2.6.1.51). This enzyme also acts as an alanine--glyoxylate aminotransferase (EC 2.6.1.44). In vertebrates, it is located in the peroxisomes and/or mitochondria. - Isopenicillin N epimerase (gene cefD). This enzyme is involved in the biosynthesis of cephalosporin antibiotics and catalyzes the reversible isomerization of isopenicillin N and penicillin N. - NifS, a protein of the nitrogen fixation operon of some bacteria and cyanobacteria. The exact function of nifS is not yet known. A highly similar protein has been found in fungi (gene NFS1 or SPL1). - The small subunit of cyanobacterial soluble hydrogenase (EC 1.12.-.-). - Hypothetical protein ycbU from Bacillus subtilis. - Hypothetical protein YFL030w from yeast. The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [LIVFYCHT]-[DGH]-[LIVMFYAC]-[LIVMFYA]-x(2)-[GSTAC]-[GSTA]- [HQR]-K-x(4,6)-G-x-[GSAT]-x-[LIVMFYSAC] [K is the pyridoxal-P attachment site]-

[1] Ouzounis C., Sander C. FEBS Lett. 322:159-164(1993).
[2] Bairoch A. Unpublished observations (1992).
[3] van der Zel A., Lam H.-M., Winkler M.E. Nucleic Acids Res. 17:8379-8379(1989).

56. Annexins repeated domain signature

Annexins [1 to 6] are a group of calcium-binding proteins that associate reversibly with membranes. They bind to phospholipid bilayers in the presence of micromolar free calcium concentration. The binding is specific for calcium and for acidic phospholipids. Annexins have been claimed to be involved in cytoskeletal interactions, phospholipase inhibition, intracellular signalling, anticoagulation, and membrane fusion. Each of these proteins consist of an N-terminal domain of variable length followed by four or eight copies of a conserved segment of sixty one residues. The repeat (sometimes known as an 'endonexin fold') consists of five alpha-helices that are wound into a right-handed superhelix [7].The proteins known to belong to the annexin family are listed below: - Annexin I (Lipocortin 1) (Calpactin 2) (p35) (Chromobindin 9). - Annexin II (Lipocortin 2) (Calpactin 1) (Protein I) (p36) (Chromobindin 8). - Annexin III (Lipocortin 3) (PAP-III). - Annexin IV (Lipocortin 4) (Endonexin I) (Protein II) (Chromobindin 4). - Annexin V (Lipocortin 5) (Endonexin 2) (VAC-alpha) (Anchorin CII) (PAP-I). - Annexin VI (Lipocortin 6) (Protein III) (Chromobindin 20) (p68) (p70). This is the only known annexin that contains 8 (instead of 4) repeats. - Annexin VII (Synexin). - Annexin VIII (Vascular anticoagulant-beta) (VAC-beta). - Annexin IX from Drosophila. - Annexin X from Drosophila. - Annexin XI (Calcyclin-associated annexin) (CAP-50). - Annexin XII from Hydra vulgaris. - Annexin XIII (Intestine-specific annexin) (ISA).The signature pattern for this domain spans positions 9 to 61 of the repeatand includes the only perfectly conserved residue (an arginine in position 22)-

Consensus pattern: [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]- x(7)-[LIVMF]-x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF]-

[1] Raynal P., Pollard H.B. Biochim. Biophys. Acta 1197:63-93(1994).

[2] Barton G.J., Newman R.H., Freemont P.S., Crumpton M.J. Eur. J. Biochem. 198:749-760(1991).

[3] Burgoyne R.D., Geisow M.J. Cell Calcium 10:1-10(1989).

[4] Haigler H.T., Fitch J.M., Jones J.M., Schlaepfer D.D. Trends Biochem. Sci. 14:48-50(1989).

[5] Klee C.B. Biochemistry 27:6645-6653(1988).

[6] Smith P.D., Moss S.E. Trends Genet. 10:241-246(1994).

[7] Huber R., Roemisch J., Paques E.-P. EMBO J. 9:3867-3874(1990).

[8] Fiedler K., Simons K. Trends Biochem. Sci. 20:177-178(1995).

57. (arf_1) ADP-ribosylation factors family signature

ADP-ribosylation factors (ARF) [1,2,3,4] are 20 Kd GTP-binding proteins involved in protein trafficking. They may modulate vesicle budding and uncoating within the Golgi apparatus. ARF's also act as allosteric activators of cholera toxin ADP-ribosyltransferase activity. They are evolutionary conserved and present in all eukaryotes. At least six forms of ARF are present in mammals and three in budding yeast. The ARF family also includes proteins highly related to ARF's but which lack the cholera toxin cofactor activity, they are collectively known as ARL's (ARF-like).ARD1 is a 64 Kd mammalian protein of unknown biological function that contains an ARF domain at its C-terminal extremity. Proteins from the ARF family are generally included in the RAS 'superfamily' of small GTP-binding proteins [5], but they are only slightly related to the other RAS proteins. They also differ from RAS proteins in that they lack cysteine residues at their C-termini and are therefore not subject to prenylation. The ARFs are N-terminally myristoylated (the ARLs have not yet been shown to be modified in such a fashion). A conserved region in the C-terminal part of ARF's and ARL's has been selected as a signature pattern.

Consensus pattern: [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)- [GSA]-[LIVMF]-x-[WK]-[LIVM]-

Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see <PDOC00017

[1] Boman A.L., Kahn R.A. Trends Biochem. Sci. 20:147-150(1995).
[2] Moss J., Vaughan M. Cell. Signal. 4.367-399(1993).
[3] Moss J., Vaughan M. Prog. Nucleic Acid Res. Mol. Biol. 45:47-65(1993).
[4] Amor J.C., Harrison D.H., Kahn R.A., Ringe D. Nature 372:704-708(1994).
[5] Valencia A., Chardin P., Wittinghofer A., Sander C. Biochemistry 30:4637-4648(1991).

(arf_2) ATP/GTP-binding site motif A (P-loop)

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is

generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted are listed below: - ATP synthase alpha and beta subunits (see <PDOC00137>). - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>). - Dynamins and dynamin-like proteins (see <PDOC00362>). - Guanylate kinase (see <PDOC00670>). - Thymidine kinase (see <PDOC00524>). - Thymidylate kinase (see <PDOC01034>). - Shikimate kinase (see <PDOC00868>). - Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>). - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>). - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran (see <PDOC00859>). - ADP-ribosylation factors family (see <PDOC00781>). - Bacterial dnaA protein (see <PDOC00771>). - Bacterial recA protein (see <PDOC00131>). - Bacterial recF protein (see <PDOC00539>). - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family (See <PDOC00388>). - Bacterial type II secretion system protein E (see <PDOC00567>).Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern: [AG]-x(4)-G-K-[ST]-

[1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).

[2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).

[3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).

[4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).

[5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).

[6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).

[7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).

[8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

58. Arginase family signatures

The following enzymes have been shown [1] to be evolutionary related: - Arginase (EC 3.5.3.1), a ubiquitous enzyme which catalyzes the degradation of arginine to ornithine and urea [2]. - Agmatinase (EC 3.5.3.11) (agmatine ureohydrolase), a prokaryotic enzyme (gene speB) that catalyzes the hydrolysis of agmatine into putrescine and urea. - Formiminoglutamase (EC 3.5.3.8) (formiminoglutamate hydrolase), a prokaryotic enzyme (gene hutG) that hydrolyzes N-formimino-glutamate into glutamate and formamide. - Hypothetical proteins from methanogenic archaebacteria. These enzymes are proteins of about 300 amino-acid residues. Three conserved regions that contain charged residues which are involved in the binding of the two manganese ions [3] can be used as signature patterns.-

Consensus pattern: [LIVMF]-G-G-x-H-x-[LIVMT]-[STAV]-x-[PAG]-x(3)-[GSTA] [H binds manganese]-

Consensus pattern: [LIVM](2)-x-[LIVMFY]-D-[AS]-H-x-D [The two D's and the H bind manganese]-

Consensus pattern: [ST]-[LIVMFY]-D-[LIVM]-D-x(3)-[PAQ]-x(3)-P-[GSA]-x(7)-G [The two D's bind manganese]

[1] Ouzounis C., Kyrpides N.C. J. Mol. Evol. 39:101-104(1994).

[2] Jenkinson C.P., Grody W.W., Cederbaum S.D. Comp. Biochem. Physiol. 114B:107-132(196).

[3] Kanyo Z.F., Scolnick L.R., Ash D.E., Christianson D.W. Nature 383:554-557(1996).

59. (asp) Eukaryotic and viral aspartyl proteases active site

Aspartyl proteases, also known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes [1,2,3] known to exist invertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue.The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases are: - Vertebrate gastric pepsins A and C (also known as gastricsin). - Vertebrate chymosin (rennin), involved in digestion and used for making cheese. - Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34). - Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma. - Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21). - Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases. - Yeast barrier pepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone. - Fission yeast sxa1 which is involved in degrading or processing the mating pheromones. Most retroviruses and some plant viruses, such as badnaviruses, encode for anaspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of apolyprotein which is cleaved during the maturation process of the virus. It is generally part of the pol polyprotein and, more rarely, of the gagpolyprotein. Conservation of the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases allows us to develop a single signature pattern for both groups of protease.

Consensus pattern: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]- x-[LIVMFSTNC]-x-[LIVMFGTA] [D is the active site residue]
Note: these proteins belong to families A1 and A2 in the classification of peptidases [4,E1

[1] Foltmann B. Essays Biochem. 17:52-84(1981).
[2] Davies D.R. Annu. Rev. Biophys. Chem. 19:189-215(1990).
[3] Rao J.K.M., Erickson J.W., Wlodawer A. Biochemistry 30:4663-4671(1991).
[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:105-120(1995).

60. (BIRA) Biotin repressor

[1] Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW; Proc Natl Acad Sci USA 1992;89:9257-9261.

61. BTB/POZ domain

The BTB (for BR-C, ttk and bab) [1] or POZ (for Pox virus and Zinc finger)[2] domain is present near the N-terminus of a fraction of zinc finger (zf-C2H2) proteins and in proteins that contain the Kelch motif such as Kelch and a family of pox virus proteins. The BTB/POZ domain mediates homomeric dimerisation and in some instances heteromeric dimerisation [2].The structure of the dimerised PLZF BTB/POZ domain has been solved and consists of a tightly intertwined homodimer. The central scaffolding of the protein is made up of a cluster of alpha-helices flanked by short beta-sheets at both the top and bottom of the molecule [3]. POZ domains from several zinc finger proteins have been shown to mediate transcriptional repression and to interact with components of histone deacetylase co-repressor complexes including N-CoR and SMRT [4,5,6]. The POZ or BTB domain is also known as BR-C/Ttk or ZiN

[1] Zollman S, Godt D, Prive GG, Couderc JL, Laski FA; Proc Natl Acad Sci U S A 1994;91:10717-10721.

[2]Bardwell VJ, Treisman R; Genes Dev 1994;8:1664-1677.

[3] Ahmad KF, Engel CK, Prive GG; Proc Natl Acad Sci U S A 1998;95:12123-12128.

[4] Deweindt C, Albagli O, Bernardin F, Dhordain P, Quief S, Lantoine D, Kerckaert JP, Leprince D; Cell Growth Differ 1995;6:1495-1503.

[5] Huynh KD, Bardwell VJ; Oncogene 1998;17:2473-2484.

[6] Wong CW, Privalsky ML; J Biol Chem 1998;273:27695-27702.

62. (Bac GSPproteins) Bacterial type II secretion system protein D signature

A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1 to 5], have been found to be evolutionary related. These proteins are listed below: - The 'D' protein from the GSP operon of: Aeromonas (gene exeD); Erwinia (gene outD); Escherichia coli (gene yheF), Klebsiella pneumoniae (gene pulD); Pseudomonas aeruginosa (gene xcpQ); Vibrio cholerae

(gene epsD) and Xanthomonas campestris (gene xpsD). - comE from Haemophilus influenzae, involved in competence (DNA uptake). - pilQ from Pseudomonas aeruginosa, which is essential for the formation of the pili. - hofQ (hopQ) from Escherichia coli. - hrpH from Pseudomonas syringae, which is involved in the secretion of a proteinaceous elicitor of the hypersensitivity response in plants. - hrpA1 from Xanthomonas campestris pv. vesicatoria, which is also involved in the hypersensitivity response. - mxiD from Shigella flexneri which is involved in the secretion of the Ipa invasins which are necessary for penetration of intestinal epithelial cells. - omc from Neisseria gonorrhoeae. - yssC from Yersinia enterocolitica virulence plasmid pYV, which seems to be required for the export of the Yop virulence proteins. - The gpIV protein from filamentous phages such as f1, ike, or m13. GpIV is said to be involved in phage assembly and morphogenesis. These proteins all seem to start with a signal sequence and are thought to be integral proteins in the outer membrane. As a signature pattern a conserved region in the C-terminal section of these proteins has been selected

Consensus pattern: [GR]-[DEQKG]-[STVM]-[LIVMA](3)-[GA]-G-[LIVMFY]-x(11)-[LIVM]-P-[LIVMFYWGS]-[LIVMF]-[GSAE]-x-[LIVM]-P- [LIVMFYW](2)-x(2)-[LV]-F

[1] Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).
[2] Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G., Bunce N., Barallon R., Douglas P., Mulholland V., Stevens S., Walker S., Salmond G.P.C. Mol. Microbiol. 8:443-456(1993).
[3] Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S. Mol. Microbiol. 9:857-868(1993).
[4] Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).
[5] Genin S., Boucher C.A. Mol. Gen. Genet. 243:112-118(1994).

63. (Bac globin) Protozoan/cyanobacterial globins signature

Globins are heme-containing proteins involved in binding and/or transporting oxygen [1]. Almost all globins belong to a large family (see <PDOC00793>), the only exceptions are the following proteins which form a family of their own[2,3]: - Monomeric hemoglobins from the protozoan Paramecium caudatum, Tetrahymena pyriformis and Tetrahymena thermophila. - Cyanoglobin from the cyanobacteria Nostoc commune. - Globins LI637 and

LI410 from the chloroplast of the alga Chlamydomonas eugametos. - Mycobacterium tuberculosis hypothetical protein MtCY48.23.These proteins contain a conserved histidine which could be involved in heme-binding. As a signature pattern, a conserved region that ends with this residue was used

Consensus pattern: F-[LF]-x(5)-G-[PA]-x(4)-G-[KRA]-x-[LIVM]-x(3)-H-

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

[2] Takagi T. Curr. Opin. Struct. Biol. 3:413-418(1993).

[3] Couture M., Chamberland H., St-Pierre B., Lafontaine J., Guertin M.; Mol. Gen. Genet. 243:185-197(1994).

64. Band 7 protein family signature

Mammalian band 7 protein [1] (also known as 7.2B or stomatin) is an integral membrane phosphoprotein of red blood cells thought to regulate cation conductance by interacting with other proteins of the junctional complex of the membrane skeleton. Structurally, band 7 is evolutionary related to the following proteins: - Caenorhabditis elegans protein mec-2 [2]. Mec-2 positively regulates the activity of the putative mechanosensory transduction channel. It may links the mechanosensory channel and the microtubule cytoskeleton of the touch receptor neurons. - Caenorhabditis elegans proteins sto-1 to sto-4. - Caenorhabditis elegans protein unc-1. - Escherichia coli hypothetical protein ybbK. - Mycobacterium tuberculosis hypothetical protein MtCY277.09. - Synechocystis strain PCC 6803 hypothetical protein slr1128. - Methanococcus jannaschii hypothetical protein MJ0827.Structurally all these proteins consist of a short N-terminal domain which is followed by a transmembrane region and a variable size (from 170 to 350residues) C-terminal domain .As a signature pattern, a conserved region located about 110residues after the transmembrane domain was selected

Consensus pattern: R-x(2)-[LIV]-[SAN]-x(6)-[LIV]-D-x(2)-T-x(2)-W-G-[LIV]- [KRH]-[LIV]-x-[KR]-[LIV]-E-[LIV]-[KR]-

[1] Gallagher P.G., Forget B.G. J. Biol. Chem. 270:26358-26363(1995).

[2] Huang M., Gu G., Ferguson E.L., Chalfie M. Nature 378:292-295(1995).

65. Barwin domain signatures

Barwin [1] is a barley seed protein of 125 residues that binds weakly a chitinanalog. It contains six cysteines involved in disulfide bonds, as shown in the following schematic representation.

+---------------+ | ***** | ****

xxxxxxxxxxxxxxxCxxxxxxxxxxCxxxxCxCxxxxxxxxCxxxxxxxxxxxxxxxxxxCx | | | | +-------
--------+ +---------------------------+'C': conserved cysteine involved in a disulfide bond.'*': position of the patterns. Barwin is closely related to the following proteins: - Hevein, a wound-induced protein found in the latex of rubber trees. - HEL, an Arabidopsis thaliana hevein-like protein [2]. - Win1 and win2, two wound-induced proteins from potato. - Pathogenesis-related protein 4 from tobacco. Hevein and the win1/2 proteins consist of an N-terminal chitin-binding domain followed by a barwin-like C-terminal domain. Barwin and its related proteins could be involved in a defense mechanism in plants. As signature patterns, two highly conserved regions that contain some of the cysteines were selected

Consensus pattern: C-G-[KR]-C-L-x-V-x-N [The two C's are involved in disulfide bonds]-
Consensus pattern: V-[DN]-Y-[EQ]-F-V-[DN]-C [C is involved in a disulfide bond]-

[1] Svensson B., Svendsen I., Hoejrup P., Roepstorff P., Ludvigsen S., Poulsen F.M. Biochemistry 31:8767-8770(1992).
[2] Potter S., Uknes S., Lawton K., Winter A.M., Chandler D., Dimaio J., Novitzky R., Ward E., Ryals J. Mol. Plant Microbe Interact. 6:680-685(1993).

66. (Bowman-Birk leg) Bowman-Birk serine protease inhibitors family signature

PROSITE cross-reference(s). The Bowman-Birk inhibitor family [1] is one of the numerous families of serine proteinase inhibitors. As it can be seen in the schematic representation, they have a duplicated structure and generally possess two distinct inhibitory sites:

```
+---------------------------------------------+
|   +-----+    +-------+   +-----+      |
|   |   |    |    |  |  |    |
```

```
xxCCxxCxxCxx#xxCxxCxxxxCxxxCxxxCxxxxCxx#xxCxxCxxCxxCxx
  ||      |********|****        ||
  ||      |     |          ||
  +--|-----------+     +----------------+ |
    +---------------------------------------+

  <-----------------70 residues------------------->
```

'C': conserved cysteine involved in a disulfide bond.
'#': active site residue.
'*': position of the pattern.

These inhibitors are found in the seeds of all leguminous plants as well as in cereal grains. In cereals they exist in two forms, one of which is a duplication of the basic structure shown above [2]. The pattern that was developed to pick up sequences belonging to this family of inhibitors is in the central part of the domain and includes four cysteines.

Consensus pattern C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C [The four C's are involved in disulfide bonds] Note this pattern can be found twice in some duplicated cereal inhibitors.

[1] Laskowski M., Kato I. Annu. Rev. Biochem. 49:593-626(1980).
[2] Tashiro M., Hashino K., Shiozaki M., Ibuki F., Maki Z. J. Biochem. 102:297-306(1987).

67. Pathogenesis-related protein Bet v I family signature

A number of plant proteins, which all seem to be involved in pathogen defense response, are structurally related [1,2,3]. These proteins are:

- Bet v I, the major pollen allergen from white birch. Bet v I is the main cause of type I allergic reactions in Europe, North America and USSR.
- Aln g I, the major pollen allergen from alder.
- Api G I, the major allergen from celery.
- Car b I, the major pollen allergen from hornbeam.

- Cor a I, the major pollen allergen from hazel.
- Mal d I, the major pollen allergen from apple.
- Asparagus wound-induced protein AoPR1.
- Kidney bean pathogenesis-related proteins 1 and 2.
- Parsley pathogenesis-related proteins PR1-1 and PR1-3.
- Pea disease resistance response proteins pI49, pI176 and DRRG49-C.
- Pea abscisic acid-responsive proteins ABR17 and ABR18.
- Potato pathogenesis-related proteins STH-2 and STH-21.
- Soybean stress-induced protein SAM22.

These proteins are thought to be intracellularly located. They contain from 155 to 160 amino acid residues. As a signature pattern, a conserved region located in the third quarter of these proteins has been selected

Consensus pattern: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]- [CS]-x(2)-K-x(4)-[FY]-

[1] Breiteneder H., Pettenburger K., Bito A., Valenta R., Kraft D., Rumpold H., Scheiner O., Breitenbach M. EMBO J. 8:1935-1938(1989).

[2] Crowell D., John M.E., Russell D., Amasino R.M. Plant Mol. Biol. 18:459-466(1992).

[3] Warner S.A.J., Scott R., Draper J. Plant Mol. Biol. 19:555-561(1992).

68. bZIP transcription factors basic domain signature

The bZIP superfamily [1,2,] of eukaryotic DNA-binding transcription factors groups together proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. This family is quite large, therefore only a parital list of some representative members appears here. - Transcription factor AP-1, which binds selectively to enhancer elements in the cis control regions of SV40 and metallothionein IIA. AP-1, also known as c-jun, is the cellular homolog of the avian sarcoma virus 17 (ASV17) oncogene v-jun. - Jun-B and jun-D, probable transcription factors which are highly similar to jun/AP-1. - The fos protein, a proto-oncogene that forms a non-covalent dimer with c-jun. - The fos-related proteins fra-1, and fos B. - Mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1. - Maize Opaque 2, a trans-acting transcriptional activator involved in the regulation of the production of zein proteins during endosperm. - Arabidopsis G-box binding factors GBF1 to GBF4,

Parsley CPRF-1 to CPRF-3, Tobacco TAF-1 and wheat EMBP-1. All these proteins bind the G-box promoter elements of many plant genes. - Drosophila protein Giant, which represses the expression of both the kruppel and knirps segmentation gap genes. - Drosophila Box B binding factor 2 (BBF-2), a transcriptional activator that binds to fat body-specific enhancers of alcohol dehydrogenase and yolk protein genes. - Drosophila segmentation protein cap'n'collar (gene cnc), which is involved in head morphogenesis. - Caenorhabditis elegans skn-1, a developmental protein involved in the fate of ventral blastomeres in the early embryo. - Yeast GCN4 transcription factor, a component of the general control system that regulates the expression of amino acid-synthesizing enzymes in response to amino acid starvation, and the related Neurospora crassa cpc-1 protein. - Neurospora crassa cys-3 which turns on the expression of structural genes which encode sulfur-catabolic enzymes. - Yeast MET28, a transcriptional activator of sulfur amino acids metabolism. - Yeast PDR4 (or YAP1), a transcriptional activator of the genes for some oxygen detoxification enzymes. - Epstein-Barr virus trans-activator protein BZLF1.-

Consensus pattern: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK]-

[1] Hurst H.C. Protein Prof. 2:105-168(1995).[2] Ellenberger T. Curr. Opin. Struct. Biol. 4:12-21(1994).

69. Biotin-requiring enzymes attachment site

Biotin, which plays a catalytic role in some carboxyl transfer reactions, is covalently attached, via an amide bond, to a lysine residue in enzymes requiring this coenzyme [1,2,3,4]. Such enzymes are:

- Pyruvate carboxylase (EC 6.4.1.1).
- Acetyl-CoA carboxylase (EC 6.4.1.2).
- Propionyl-CoA carboxylase (EC 6.4.1.3).
- Methylcrotonoyl-CoA carboxylase (EC 6.4.1.4).
- Geranoyl-CoA carboxylase (EC 6.4.1.5).
- Urea carboxylase (EC 6.3.4.6).
- Oxaloacetate decarboxylase (EC 4.1.1.3).
- Methylmalonyl-CoA decarboxylase (EC 4.1.1.41).
- Glutaconyl-CoA decarboxylase (EC 4.1.1.70).

- Methylmalonyl-CoA carboxyl-transferase (EC 2.1.3.1) (transcarboxylase).

Sequence data reveal that the region around the biocytin (biotin-lysine) residue is well conserved and can be used as a signature pattern.

Consensus pattern[GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-[SAV] [K is the biotin attachment site] Note the domain around the biotin-binding lysine residue is evolutionary related to that around the lipoyl-binding lysine residue of 2-oxo acid dehydrogenase acyltransferases

[1] Knowles J.R. Annu. Rev. Biochem. 58:195-221(1989).
[2] Samols D., Thronton C.G., Murtif V.L., Kumar G.K., Haase F.C., Wood H.G. J. Biol. Chem. 263:6461-6464(1988).
[3] Goss N.H., Wood H.G. Meth. Enzymol. 107:261-278(1984).
[4] Shenoy B.C., Xie Y., Park V.L., Kumar G.K., Beegen H., Wood H.G., Samols D. J. Biol. Chem. 267:18407-18412(1992).

2-oxo acid dehydrogenases acyltransferase component lipoyl binding site

The 2-oxo acid dehydrogenase multienzyme complexes [1,2] from bacterial and eukaryotic sources catalyze the oxidative decarboxylation of 2-oxo acids to the corresponding acyl-CoA. The three members of this family of multienzyme complexes are:

- Pyruvate dehydrogenase complex (PDC).
- 2-oxoglutarate dehydrogenase complex (OGDC).
- Branched-chain 2-oxo acid dehydrogenase complex (BCOADC).

These three complexes share a common architecture: they are composed of multiple copies of three component enzymes - E1, E2 and E3. E1 is a thiamine pyrophosphate-dependent 2-oxo acid dehydrogenase, E2 a dihydrolipamide acyltransferase, and E3 an FAD-containing dihydrolipamide dehydrogenase.

E2 acyltransferases have an essential cofactor, lipoic acid, which is covalently bound via a amide linkage to a lysine group. The E2 components of OGCD and BCOACD bind a single lipoyl group, while those of PDC bind either one (in yeast and in Bacillus), two (in mammals), or three (in Azotobacter and in Escherichia coli) lipoyl groups [3].

In addition to the E2 components of the three enzymatic complexes described

above, a lipoic acid cofactor is also found in the following proteins:

- H-protein of the glycine cleavage system (GCS) [4]. GCS is a multienzyme complex of four protein components, which catalyzes the degradation of glycine. H protein shuttles the methylamine group of glycine from the P protein to the T protein. H-protein from either prokaryotes or eukaryotes binds a single lipoic group.
- Mammalian and yeast pyruvate dehydrogenase complexes differ from that of other sources, in that they contain, in small amounts, a protein of unknown function - designated protein X or component X. Its sequence is closely related to that of E2 subunits and seems to bind a lipoic group [5].
- Fast migrating protein (FMP) (gene acoC) from Alcaligenes eutrophus [6]. This protein is most probably a dihydrolipamide acyltransferase involved in acetoin metabolism.

A signature pattern was developed which allows the detection of the lipoyl-binding site.

Consensus pattern[GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-[STAVQDN]-x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY] [K is the lipoyl-binding site] Note the domain around the lipoyl-binding lysine residue is evolutionary related to that around the biotin-binding lysine residue of biotin requiring enzymes

[1] Yeaman S.J. Biochem. J. 257:625-632(1989).

[2] Yeaman S.J. Trends Biochem. Sci. 11:293-296(1986).

[3] Russel G.C., Guest J.R. Biochim. Biophys. Acta 1076:225-232(1991).

[4] Fujiwara K., Okamura-Ikeda K., Motokawa Y. J. Biol. Chem. 261:8836-8841(1986).

[5] Behal R.H., Browning K.S., Hall T.B., Reed L.J. Proc. Natl. Acad. Sci. U.S.A. 86:8732-8736(1989).

[6] Priefert H., Hein S., Krueger N., Zeh K., Schmidt B., Steinbuechel A. J. Bacteriol. 173:4056-4071(1991).

70. C2 (C2 domain) Number of members: 295

Some isozymes of protein kinase C (PKC) [1,2] contain a domain, known as C2, of about 116 amino-acid residues which is located between the two copies of the C1 domain (that

bind phorbol esters and diacylglycerol) (see <PDOC00379>) and the protein kinase catalytic domain (see <PDOC00100>). Regions with significant homology [3,E1] to the C2-domain have been found in the following proteins:

- PKC isoforms alpha, beta and gamma and Drosophila isoforms PKC1 and PKC2.
- PKC isoforms delta, epsilon and eta, Caenorhabditis elegans kin-13 and yeast PKC1 have a C2-like domain at the N-terminal extremity [4].
- Yeast cAMP dependent protein kinase SCH9 contains a C2-like domain.
- Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) (see <PDOC50007>) isoforms beta, gamma and delta as well as several non-mammalian PI-PLCs have a C2-like domain C-terminal of the catalytic domain.
- Mammalian and plants phosphatidylinositol-3-kinase have a C2-like domain in the central region of the 110 Kd catalytic subunit.
- Yeast phosphatidylserine-decarboxylase 2 (gene PSD2) contains a C2 domain in its central region.
- Cytosolic phospholipase D from plants and cytosolic phospholipase A2 have a C2-like domain at their N-terminus.
- Synaptotagmins (p65). This is a family of related synaptic vesicle proteins that bind acidic phospholipids and that may have a regulatory role in the membrane interactions during trafficking of synaptic vesicles at the active zone of the synapse. All isoforms of synaptotagmins have two copies of the C2 domain in their C-terminal region.
- Rabphilin-3A, a synaptic protein contains two C2 domains.
- Caenorhabditis elegans protein unc-13 whose function is not known. Unc-13 has a C2 domain in its central part and a C2-like domain at the C-terminus.
- rasGAP and the breakpoint cluster protein bcr have a C2-domain C-terminal of a PH-domain.
- Yeast protein BUD2 (or CLA2) has a C2-domain in the central region.
- Yeast protein RSP5 and human protein NEDD-4, both proteins also contain WW domains (see <PDOC50020>).
- Perforin (see <PDOC00251>) has a C2 domain at the C-terminus. It is the only extracellular protein known to contain a C2 domain.
- Yeast hypothetical protein YML072C has a C2 domain.
- Yeast hypothetical protein YNL087W has three C2 domains.
- Caenorhabditis elegans hypothetical protein F37A4.7 has two C2 domains.

The C2 domain is thought to be involved in calcium-dependent phospholipid binding [5]. Since domains related to the C2 domain are also found in proteins that do not bind calcium, other putative functions for the C2 domain like e.g. binding to inositol-1,3,4,5-tetraphosphate have been suggested [6]. Recently, the 3D structure of the first C2 domain of synaptotagmin has been reported [7], the domain forms an eight-stranded beta sandwich. The signature pattern that has been developed for the C2 domain is located in a conserved part of that domain, the connecting loop between beta strands 2 and 3. A profile has been developed for the C2 domain that covers the total domain.

-Consensus pattern: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D-[PA]-[FY]

-Note: this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

[1]Medline: 96367095 Extending the C2 domain family: C2s in PKCs delta, epsilon, eta and theta, phospholipases, GAPs and perforin. Ponting CP, Parker PJ; Protein Sci 1996;5:162-166.

[1] Azzi A., Boscoboinik D., Hensey C. Eur. J. Biochem. 208:547-557(1992).

[2] Stabel S. Semin. Cancer Biol. 5:277-284(1994).

[3] Brose N., Hofmann K.O., Hata Y., Suedhof T.C. J. Biol. Chem. 270:25273-25280(1995).

[4] Sossin W.S., Schwartz J.H. Trends Biochem. Sci. 18:207-208(1993).

[5] Davletov B.A., Suedhof T.C. J. Biol. Chem. 268:26386-26390(1993).

[6] Fukuda M., Aruga J., Niinobe M., Aimoto S., Mikoshiba K. J. Biol. Chem. 269:29206-29211(1994).

[6] Sutton R.B., Davletov B.A., Berghuis A.M., Suedhof T.C., Sprang S.R. Cell 80:929-938(1995).

71. CAP (CAP protein) Number of members: 11

In budding and fission yeasts the CAP protein is a bifunctional protein whose N-terminal domain binds to adenylyl cyclase, thereby enabling that enzyme to be activated by upstream regulatory signals, such as Ras. The function of the C-terminal domain is less clear, but it is

required for normal cellular morphology and growth control [1]. CAP is conserved in higher eukaryotic organisms where its function is not yet clear [2].

Structurally, CAP is a protein of 474 to 551 residues which consist of two domains separated by a proline-rich hinge. Two signature patterns, one corresponding to a conserved region in the N-terminal extremity and the other to a C-terminal region have been developed.

-Consensus pattern: [LIVM](2)-x-R-L-[DE]-x(4)-R-L-E

-Consensus pattern: D-[LIVMFY]-x-E-x-[PA]-x-P-E-Q-[LIVMFY]-K

[1] Kawamukai M., Gerst J., Field J., Riggs M., Rodgers L., Wigler M., Young D. Mol. Biol. Cell 3:167-180(1992).

[2] Yu G., Swiston J., Young D. J. Cell Sci. 107:1671-1678(1994).

72. CAP_GLY (CAP-Gly domain)

CAP stands for cytoskeleton-associated proteins. Swiss:P39937 may be a member but has not been included. It has a weak match to the family between residues 22-67. Number of members: 24

[1]Medline: 93242656. Sequence homologies between four cytoskeleton-associated proteins. Riehemann K, Sorg C; Trends Biochem Sci 1993;18:82-83.

It has been shown [1] that some cytoskeleton-associated proteins (CAP) share the presence of a conserved, glycine-rich domain of about 42 residues, called here CAP-Gly. Proteins known to contain this domain are listed below.

- Restin (also known as cytoplasmic linker protein-170 or CLIP-170), a 160 Kd protein associated with intermediate filaments and that links endocytic vesicles to microtubules. Restin contains two copies of the CAP-Gly domain.
- Vertebrate dynactin (150 Kd dynein-associated polypeptide; DAP) and Drosophila glued, a major component of activator I, a 20S polypeptide complex that stimulates dynein-mediated vesicle transport.
- Yeast protein BIK1 which seems to be required for the formation or stabilization of microtubules during mitosis and for spindle pole body fusion during conjugation.
- Yeast protein NIP100 (NIP80).

- Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain a N-terminal ubiquitin domain (see <PDOC00271>) and a C-terminal CAP-Gly domain.
- Caenorhabditis elegans hypothetical protein M01A8.2.
- Yeast hypothetical protein YNL148c.

Structurally, these proteins are made of three distinct parts: an N-terminal section that is most probably globular and contains the CAP-Gly domain, a large central region predicted to be in an alpha-helical coiled-coil conformation and, finally, a short C-terminal globular domain. The signature for the CAP-Gly domain corresponds to the first 32 residues of the domain and includes five of the six conserved glycines.

-Consensus pattern: G-x(8,10)-[FYW]-x-G-[LIVM]-x-[LIVMFY]-x(4)-G-K-[NH]-x-G-[STAR]-x(2)-G-x(2)-[LY]-F

[1] Riehemann K., Sorg C. Trends Biochem. Sci. 18:82-83(1993).

73. (CBD 1)

Cellulose-binding domain, fungal type

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1].

Structurally, cellulases and xylanases generally consist of a catalytic domain joined to a cellulose-binding domain (CBD) by a short linker sequence rich in proline and/or hydroxy-amino acids.

The CBD of a number of fungal cellulases has been shown to consist of 36 amino acid residues. Enzymes known to contain such a domain are:

- Endoglucanase I (gene egl1) from Trichoderma reesei.
- Endoglucanase II (gene egl2) from Trichoderma reesei.
- Endoglucanase V (gene egl5) from Trichoderma reesei.
- Exocellobiohydrolase I (gene CBHI) from Humicola grisea, Neurospora crassa,

Phanerochaete chrysosporium, Trichoderma reesei, and Trichoderma viride.

- Exocellobiohydrolase II (gene CBHII) from Trichoderma reesei.
- Exocellobiohydrolase 3 (gene cel3) from Agaricus bisporus
- Endoglucanases B, C2, F and K from Fusarium oxysporum.

The CBD domain is found either at the N-terminal (Cbh-II or egl2) or at the C-terminal extremity (Cbh-I, egl1 or egl5) of these enzymes. As it is shown in the following schematic representation, there are four conserved cysteines in this type of CBD domain, all involved in disulfide bonds.

```
      +----------------+
      |      +-----|---------+
      |      |   |      |
xxxxxxxCxxxxxxxxxxCxxxxxCxxxxxxxxxCx
      ****************************
```

'C': conserved cysteine involved in a disulfide bond.
'*': position of the pattern.

Such a domain has also been found in a putative polysaccharide binding protein from the red alga, Porphyra purpurea [2]. Structurally, this protein consists of four tandem repeats of the CBD domain.

Consensus patternC-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]- x(2)-Q-C [The four C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern ALL.

[1] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).
[2] Liu Q., der Meer J.P., Reith M.E.

74. CBS domain. 3D Structure found as a subdomain in TIM barrel of inosine-. CBS domain web page. CBS domains are small intracellular modules mostly found in 2 or four copies

within a protein. CBS domains are found in cystathionine-beta-synthase (CBS) where mutations lead to homocystinuria. Two CBS domains are found in inosine-monophosphate dehydrogenase from all species, however the CBS domains are not needed for activity. Two CBS domains are found in intracellular loops of several chloride channels. Mutations in this domain of Swiss:P35520 lead to homocystinuria.

Number of members: 414

[1]Medline: 97172695 The structure of a domain common to archaebacteria and the homocystinuria disease protein. Bateman A; Trends Biochem Sci 1997;22:12-13.

[2]Medline: 96279836 Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic-acid. Sintchak MD, Fleming MA, Futer O, Raybuck SA, Chambers SP, Caron PR, Murcko MA, Wilson KP; Cell 1996;85:921-930.

Discovery of CBS domain.

[3]Medline: 97259972 CBS domains in ClC chloride channels implicated in myotonia and nephrolithiasis (kidney stones). Ponting CP; J Mol Med 1997;75:160-163.

75. CDP-OH_P_transf (CDP-alcohol phosphatidyltransferase)

All of these members have the ability to catalyze the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond. Number of members: 32

A number of phosphatidyltransferases, which are all involved in phospholipid biosynthesis and that share the property of catalyzing the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond share a conserved sequence region [1,2]. These enzymes are:

- Ethanolaminephosphotransferase (EC 2.7.8.1) from yeast (gene EPT1).
- Diacylglycerol cholinephosphotransferase (EC 2.7.8.2) from yeast (gene CPT1).
- Phosphatidylglycerophosphate synthase (EC 2.7.8.5) (CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase) from bacteria (gene pgsA).
- Phosphatidylserine synthase (EC 2.7.8.8) (CDP-diacylglycerol--serine O-phosphatidyltransferase) from yeast (gene CHO1) and from Bacillus subtilis (gene pssA).
- Phosphatidylinositol synthase (EC 2.7.8.11) (CDP-diacylglycerol--inositol 3-phosphatidyltransferase) from yeast (gene PIS).

These enzymes are proteins of from 200 to 400 amino acid residues. The conserved region contains three aspartic acid residues and is located in the N-terminal section of the sequences.

-Consensus pattern: D-G-x(2)-A-R-x(8)-G-x(3)-D-x(3)-D

[1]Medline: 97075020 Two-dimensional 1H-NMR of transmembrane peptides from Escherichia coli phosphatidylglycerophosphate synthase in micelles. Morein S, Trouard TP, Hauksson JB, Rilfors L, Arvidson G, Lindblom G; Eur J Biochem 1996;241:489-497.

[1] Nikawa J.-I., Kodaki T., Yamashita S.
J. Biol. Chem. 262:4876-4881(1987).
[2] Hjelmstad R.H., Bell R.M.
J. Biol. Chem. 266:5094-5134(1991).

76. CHOD (Cholesterol oxidase) Members of the GMC oxidoreductase family. Number of members: 3

[1]Medline: 94032271. Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. Li J, Vrielink A, Brick P, Blow DM; Biochemistry 1993;32:11507-11515.

The following FAD flavoproteins oxidoreductases have been found [1,2] to be evolutionary related. These enzymes, which are called 'GMC oxidoreductases', are listed below.

- Glucose oxidase (EC 1.1.3.4) (GOX) from Aspergillus niger. Reaction catalyzed: glucose + oxygen -> delta-luconolactone + hydrogen peroxide.
- Methanol oxidase (EC 1.1.3.13) (MOX) from fungi. Reaction catalyzed: methanol + oxygen -> acetaldehyde + hydrogen peroxide.
- Choline dehydrogenase (EC 1.1.99.1) (CHD) from bacteria. Reaction catalyzed: choline + unknown acceptor -> betaine acetaldehyde + reduced acceptor.
- Glucose dehydrogenase (GLD) (EC 1.1.99.10) from Drosophila. Reaction catalyzed: glucose + unknown acceptor -> delta-gluconolactone + reduced acceptor.

- Cholesterol oxidase (CHOD) (EC 1.1.3.6) from Brevibacterium sterolicum and Streptomyces strain SA-COO. Reaction catalyzed: cholesterol + oxygen -> cholest-4-en-3-one + hydrogen peroxide.

- AlkJ [3], an alcohol dehydrogenase from Pseudomonas oleovorans, which converts aliphatic medium-chain-length alcohols into aldehydes. This family also includes a lyase:

- (R)-mandelonitrile lyase (EC 4.1.2.10) (hydroxynitrile lyase) from plants [4], an enzyme involved in cyanogenis, the release of hydrogen cyanide from injured tissues.

These enzymes are proteins of size ranging from 556 (CHD) to 664 (MOX) amino acid residues which share a number of regions of sequence similarities. One of these regions, located in the N-terminal section, corresponds to the FAD ADP- binding domain. The function of the other conserved domains is not yet known; two of these domains have been selected as signature patterns. The first one is located in the N-terminal section of these enzymes, about 50 residues after the ADP-binding domain, while the second one is located in the central section.

-Consensus pattern: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]- x(2)-[PAG]-x(5)-[DNESH]

-Consensus pattern: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G

[1] Cavener D.R. J. Mol. Biol. 223:811-814(1992).

[2] Henikoff S., Henikoff J.G. Genomics 19:97-107(1994).

[3] van Beilen J.B., Eggink G., Enequist H., Bos R., Witholt B. Mol. Microbiol. 6:3121-3136(1992).

[4] Cheng I.P., Poulton J.E. Plant Cell Physiol. 34:1139-1143(1993).

77. CKS (Cyclin-dependent kinase regulatory subunit) Number of members: 11. Cyclin-dependent kinases (CDK) are protein kinases which associate with cyclins to regulate eukaryotic cell cycle progression. The most well known CDK is p34-cdc2 (CDC28 in yeast) which is required for entry into S-phase and mitosis. CDK's bind to a regulatory subunit which is essential for their biological function. This regulatory subunit is a small protein of 79 to 150 residues. In yeast (gene CKS1) and in fission yeast (gene suc1) a single isoform is known, while mammals have two highly related isoforms. It has been shown [1] that these CDK regulatory subunits assemble as an hexamer which then acts as a hub for the

oligomerization of six CDK catalytic subunits. The sequence of CDK regulatory subunits are highly conserved therefore, the two most conserved regions have been used as signature patterns.

-Consensus pattern: Y-S-x-[KR]-Y-x-[DE](2)-x-[FY]-E-Y-R-H-V-x-[LV]-[PT]-[KRP]
-Consensus pattern: H-x-P-E-x-H-[IV]-L-L-F-[KR]

[1] Parge H.E., Arvai A.S., Murtari D.J., Reed S.I., Tainer J.A. Science 262:387-395(1993).

78. CK_II_beta (Casein kinase II regulatory subunit)

Number of members: 16. Casein kinase II (CK-2) [1] is an ubiquitous eukaryotic serine/threonine protein kinase which is found both in the cytoplasm and the nucleus and whose substrates are numerous. It generally phosphorylates Ser or Thr at the N-terminal of stretch of acidic residues (see <PDOC00006>). CK-2 exists as an heterotetramer composed of two catalytic subunits (alpha) and two regulatory subunits (beta). In most species there are two closely related isoforms of the catalytic subunit: alpha and alpha'. Some species, such as fungi and plants, express two forms of regulatory subunits: beta and beta'. The exact function of the regulatory subunit is not yet known. It is a highly conserved protein of about 25 Kd that contains, in its central section, a cysteine-rich motif that could be involved in binding a metal such as zinc [2]. This region has been used as a signature pattern.

-Consensus pattern: C-P-x-[LIVMY]-x-C-x(5)-[LI]-P-[LIVMC]-G-x(9)-V-[KR]-x(2)-C-P-x-C

[1] Allende J.E., Allende C.C. FASEB J. 9:313-323(1995).
[2] Reed J.C., Bidwai A.P., Glover C.V.C. J. Biol. Chem. 269:18192-18200(1994).

79. CLP_protease (Clp protease)

These proteins belong to family S14 in the classification of peptidases.

-!- The Clp protease has an active site catalytic triad. In E. coli Clp protease, ser-111, his-136 and asp-185 form the catalytic triad.

-!- Swiss:P48254 has lost all of these active site residues and is therefore inactive.

-!- Swiss:P42379 contains two large insertions, Swiss:P42380 contains one large insertion.

Number of members: 38

The endopeptidase Clp (EC 3.4.21.92) from Escherichia coli cleaves peptides in various proteins in a process that requires ATP hydrolysis [1,2]. Clp is a dimeric protein which consists of a proteolytic subunit (gene clpP) and either of two related ATP-binding regulatory subunits (genes clpA and clpX). ClpP is a serine protease which has a chymotrypsin-like activity. Its catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. Proteases highly similar to ClpP have been found to be encoded in the genome of the chloroplast of plants and seem to be also present in other eukaryotes. The sequences around two of the residues involved in the catalytic triad (a serine and a histidine) are highly conserved and can be used as signature patterns specific to that category of proteases.

-Consensus pattern: T-x(2)-[LIVMF]-G-x-A-[SAC]-S-[MSA]-[PAG]-[STA] [S is the active site residue]

-Consensus pattern: R-x(3)-[EAP]-x(3)-[LIVMFYT]-M-[LIVM]-H-Q-P [H is the active site residue]

[1]Medline: 98050920. The structure of ClpP at 2.3 angstroms resolution suggests a model for ATP-dependent proteolysis. Wang J, Hartling JA, Flanagan JM; Cell 1997;91:447-456.

[1] Maurizi M.R., Clark W.P., Kim S.-H., Gottesman S. J. Biol. Chem. 265:12546-12552(1990).

[2] Gottesman S., Maurizi M.R. Microbiol. Rev. 56:592-621(1992).

[3] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

80. CNG_membrane (Transmembrane region cyclic Nucleotide Gated Channel)

[1]Medline: 94224763. Cyclic nucleotide-gated channels: an expanding new family of ion channels. Yau KW; Proc Natl Acad Sci USA 1994;91:3481-3483.

This family is found to the N-terminus of the cNMP_binding. Number of members: 56.

Proteins that bind cyclic nucleotides (cAMP or cGMP) share a structural domain of about 120 residues [1-3]. The best studied of these proteins is the prokaryotic catabolite gene

activator (also known as the cAMP receptor protein) (gene crp) where such a domain is known to be composed of three alpha-helices and a distinctive eight-stranded, antiparallel beta-barrel structure. Such a domain is known to exist in the following proteins:

- Prokaryotic catabolite gene activator protein (CAP).
- cAMP- and cGMP-dependent protein kinases (cAPK and cGPK). Both types of kinases contains two tandem copies of the cyclic nucleotide-binding domain. The cAPK's are composed of two different subunits: a catalytic chain and a regulatory chain which contains both copies of the domain. The cGPK's are single chain enzymes that include the two copies of the domain in their N-terminal section. The nucleotide specificity of cAPK and cGPK is due to an amino acid in the conserved region of beta-barrel 7: a threonine that is invariant in cGPK is an alanine in most cAPK.
- Vertebrate cyclic nucleotide-gated ion-channels. Two such cations channels have been fully characterized. One is found in rod cells where it plays a role in visual signal transduction. It specifically binds to cGMP leading to an opening of the channel and thereby causing a depolarization of rod photoreceptors. In olfactory epithelium a similar, cAMP-binding, channel plays a role in odorant signal transduction. There are six invariant amino acids in this domain, three of which are glycine residues that are thought to be essential for maintenance of the structural integrity of the beta-barrel. Two signature patterns have been developed for this domain. The first pattern is located within beta-barrels and 3 and contains the first two conserved Gly. The second pattern is located within beta-barrels 6 and 7 and contains the third conserved Gly as well as the three other invariant residues.

-Consensus pattern: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G
-Consensus pattern: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV]

[1] Weber I.T., Shabb J.B., Corbin J.D. Biochemistry 28:6122-6127(1989).
[2] Kaupp U.B. Trends Neurosci. 14:150-157(1991).
[3] Shabb J.B., Corbin J.D. J. Biol. Chem. 267:5723-5726(1992).

81. COX10_ctaB_cyoE (Cytochrome c oxidase assembly factor)
[1]Medline: 95191390

Biosynthesis and functional role of haem O and haem A

Mogi T, Saiki K, Anraku Y; Mol Microbiol 1994;14:391-398.

Cytochrome c oxidase is a multi subunit enzyme. The complexity of this enzyme requires assistance in building the complex. This is carried out by the Cytochrome c oxidase assembly factor.

Number of members: 31

Cytochrome c oxidase is an oligomeric enzymatic complex which seems to require the aid of a number of proteins that either act as chaperonins to help the subunits of the enzyme to fold correctly, or assist in the assembly of the metal centers [1]. One of these subunits is known as COX10 in yeast and as ctaB [2] in aerobic prokaryotes. It is evolutionary related to cyoE protein from the Escherichia coli cytochrome O terminal oxidase complex.

These proteins probably contain [3] seven transmembrane segments. The most conserved region is located in a loop between the second and third of these segments and has been selected as a signature pattern.

-Consensus pattern: [ED]-x-D-x(2)-M-x-R-T-x(2)-R-x(4)-G

[1] Nobrega M.P., Nobrega F.G., Tzagoloff A.
J. Biol. Chem. 265:14220-14226(1990).

[2] Cao J., Hosler J., Shapleigh J., Revzin A., Ferguson-Miller S.
J. Biol. Chem. 267:24273-24278(1992).

[3] Chepuri V., Gennis R.B.
J. Biol. Chem. 265:12978-12986(1990).

82. COX3 (Cytochrome c oxidase subunit III)

This family corresponds to chains c and p.

[1]Medline: 96216288

The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S; Science 1996;272:1136-1144.

Number of members: 224

83. COX5B (Cytochrome c oxidase subunit Vb)

[1]

Medline: 96216288

The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S;

Science 1996;272:1136-1144.

This family consists of chains F and S

Number of members: 10

Cytochrome c oxidase (EC 1.9.3.1) [1] is an oligomeric enzymatic complex which is a component of the respiratory chain complex and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. In addition to the three large subunits that form the catalytic center of the enzyme complex there are, in eukaryotes, a variable number of small polypeptidic subunits. One of these subunits which is known as Vb in mammals, V in slime mold and IV in yeast, binds a zinc atom. The sequence of subunit Vb is well conserved and includes three conserved cysteines that are thought to coordinate the zinc ion [2]. Two of these cysteines are clustered in the C-terminal section of the subunit; this region has been selected as a signature pattern.

-Consensus pattern: [LIVM](2)-[FYW]-x(10)-C-x(2)-C-G-x(2)-[FY]-K-L [The two C's probably bind zinc]

[1] Capaldi R.A., Malatesta F., Darley-Usmar V.M.
Biochim. Biophys. Acta 726:135-148(1983).

[2] Rizzuto R., Sandona D., Brini M., Capaldi R.A., Bisson R.
Biochim. Biophys. Acta 1129:100-104(1991).

84. COesterase (Carboxylesterases)

Cholinesterase pages

The prints entry is specific to acetylcholinesterase

Number of members: 273

Higher eukaryotes have many distinct esterases. Among the different types are those which act on carboxylic esters (EC 3.1.1.-). Carboxyl-esterases have been classified into three categories (A, B and C) on the basis of differential patterns of inhibition by organophosphates. The sequence of a number of type-B carboxylesterases indicates [1,2,3] that the majority are evolutionary related. This family currently consists of the following proteins:

- Acetylcholinesterase (EC 3.1.1.7) (AChE) [E1] from vertebrates and from Drosophila.
- Mammalian cholinesterase II (butyryl cholinesterase) (EC 3.1.1.8). Acetylcholinesterase and cholinesterase II are closely related enzymes that hydrolyze choline esters [4].
- Mammalian liver microsomal carboxylesterases (EC 3.1.1.1).
- Drosophila esterase 6, produced in the anterior ejaculatory duct of the male insect reproductive system where it plays an important role in its reproductive biology.
- Drosophila esterase P.
- Culex pipiens (mosquito) esterases B1 and B2.
- Myzus persicae (peach-potato aphid) esterases E4 and FE4.
- Mammalian bile-salt-activated lipase (BAL) [5], a multifunctional lipase which catalyzes fat and vitamin absorption. It is activated by bile salts in infant intestine where it helps to digest milk fats.
- Insect juvenile hormone esterase (JH esterase) (EC 3.1.1.59).
- Lipases (EC 3.1.1.3) from the fungi Geotrichum candidum and Candida rugosa.
- Caenorhabditis gut esterase (gene ges-1).
- Duck fatty acyl-CoA hydrolase, medium chain (EC 3.1.2.14), an enzyme that

may be associated with peroxisome proliferation and may play a role in the production of 3-hydroxy fatty acid diester pheromones.

- Membrane enclosed crystal proteins from slime mold. These proteins are, most probably esterases; the vesicles where they are found have therefore been termed esterosomes.

So far two bacterial proteins have been found to belong to this family:

- Phenmedipham hydrolase (phenylcarbamate hydrolase), an Arthrobacter oxidans plasmid-encoded enzyme (gene pcd) that degrades the phenylcarbamate herbicides phenmedipham and desmedipham by hydrolyzing their central carbamate linkages.
- Para-nitrobenzyl esterase from Bacillus subtilis (gene pnbA).

The following proteins, while having lost their catalytic activity, contain a domain evolutionary related to that of carboxylesterases type-B:

- Thyroglobulin (TG), a glycoprotein specific to the thyroid gland, which is the precursor of the iodinated thyroid hormones thyroxine (T4) and triiodo thyronine (T3).
- Drosophila protein neuractin (gene nrt) which may mediate or modulate cell adhesion between embryonic cells during development.
- Drosophila protein glutactin (gene glt), whose function is not known.

As is the case for lipases and serine proteases, the catalytic apparatus of esterases involves three residues (catalytic triad): a serine, a glutamate or aspartate and a histidine. The sequence around the active site serine is well conserved and can be used as a signature pattern. A conserved region located in the N-terminal section containing a cysteine involved in a disulfide bond has been selected as a second signature pattern.

-Consensus pattern: F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G[S is the active site residue]

-Consensus pattern: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR] [C is involved in a disulfide bond]

[1] Myers M., Richmond R.C., Oakeshott J.G. Mol. Biol. Evol. 5:113-119(1988).

[2] Krejci E., Duval N., Chatonnet A., Vincens P., Massoulie J. Proc. Natl. Acad. Sci. U.S.A. 88:6647-6651(1991).

[3] Cygler M., Schrag J.D., Sussman J.L., Harel M., Silman I. Gentry M.K., Doctor B.P. Protein Sci. 2:366-382(1993).

[4] Lockridge O. BioEssays 9:125-128(1988).

[5] Wang C.-S., Hartsuck J.A. Biochim. Biophys. Acta 1166:1-19(1993).

85. CPSase_L_chain (Carbamoyl-phosphate synthase (CPSase))

[1]

Medline: 94347758

Three-dimensional structure of the biotin carboxylase subunit. of acetyl-CoA carboxylase.

Waldrop GL, Rayment I, Holden HM;

Biochemistry 1994;33:10249-10256.

[1]

Medline: 90285162

Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD.

Simmer JP, Kelly RE, Rinker AG Jr, Scully JL, Evans DR;

Biol Chem 1990;265:10395-10402.

Carbamoyl-phosphate synthase catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine or ammonia and bicarbonate. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and/or pyrimidines [2].

The carbamoyl-phosphate synthase (CPS) enzyme in prokaryotes is a heterodimer of a small and large chain. The small chain promotes the hydrolysis of glutamine to ammonia, which is used by the large chain to synthesize carbamoyl phosphate. See CPSase_sm_chain.

The small chain has a GATase domain in the carboxyl terminus.
See GATase.
Number of members: 181

Carbamoyl-phosphate synthase (CPSase) catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine (EC 6.3.5.5) or ammonia (EC 6.3.4.16) and bicarbonate [1]. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and pyrimidines.

Glutamine-dependent CPSase (CPSase II) is involved in the biosynthesis of pyrimidines and purines. In bacteria such as Escherichia coli, a single enzyme is involved in both biosynthetic pathways while other bacteria have separate enzymes. The bacterial enzymes are formed of two subunits. A small chain (gene carA) that provides glutamine amidotransferase activity (GATase) necessary for removal of the ammonia group from glutamine, and a large chain (gene carB) that provides CPSase activity. Such a structure is also present in fungi for arginine biosynthesis (genes CPA1 and CPA2). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme - called URA2 in yeast, rudimentary in Drosophila and CAD in mammals [2]. The CPSase domain is located between an N-terminal GATase domain and the C-terminal part which encompass the dihydroorotase and aspartate transcarbamylase activities.

Ammonia-dependent CPSase (CPSase I) is involved in the urea cycle in ureolytic vertebrates; it is a monofunctional protein located in the mitochondrial matrix.

The CPSase domain is typically 120 Kd in size and has arisen from the duplication of an ancestral subdomain of about 500 amino acids. Each subdomain independently binds to ATP and it is suggested that the two homologous halves act separately, one to catalyze the phosphorylation of bicarbonate to carboxy phosphate and the other that of carbamate to carbamyl phosphate.

The CPSase subdomain is also present in a single copy in the biotin-dependent

enzymes acetyl-CoA carboxylase (EC 6.4.1.2) (ACC), propionyl-CoA carboxylase (EC 6.4.1.3) (PCCase), pyruvate carboxylase (EC 6.4.1.1) (PC) and urea carboxylase (EC 6.3.4.6).

Two conserved regions which are probably important for binding ATP and/or catalytic activity have been selected as signatures for the subdomain.

-Consensus pattern: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG]

-Consensus pattern: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC]

[1] Simmer J.P., Kelly R.E., Rinker A.G. Jr., Scully J.L., Evans D.R. J. Biol. Chem. 265:10395-10402(1990).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

86. CPSase_sm_chain (Carbamoyl-phosphate synthase small chain, CPSase domain) [1]

Medline: 90285162

Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD.

Simmer JP, Kelly RE, Rinker AG Jr, Scully JL, Evans DR;

Biol Chem 1990;265:10395-10402.

The carbamoyl-phosphate synthase domain is in the amino terminus of protein.

Carbamoyl-phosphate synthase catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine or ammonia and bicarbonate. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and/or pyrimidines [1].

The carbamoyl-phosphate synthase (CPS) enzyme in prokaryotes is a heterodimer of a small and large chain. The small chain promotes the hydrolysis of glutamine to ammonia, which is used by the large

chain to synthesize carbamoyl phosphate. See CPSase_L_chain. The small chain has a GATase domain in the carboxyl terminus. See GATase.

Number of members: 46

Carbamoyl-phosphate synthase (CPSase) catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine (EC 6.3.5.5) or ammonia (EC 6.3.4.16) and bicarbonate [1]. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and pyrimidines.

Glutamine-dependent CPSase (CPSase II) is involved in the biosynthesis of pyrimidines and purines. In bacteria such as Escherichia coli, a single enzyme is involved in both biosynthetic pathways while other bacteria have separate enzymes. The bacterial enzymes are formed of two subunits. A small chain (gene carA) that provides glutamine amidotransferase activity (GATase) necessary for removal of the ammonia group from glutamine, and a large chain (gene carB) that provides CPSase activity. Such a structure is also present in fungi for arginine biosynthesis (genes CPA1 and CPA2). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme - called URA2 in yeast, rudimentary in Drosophila and CAD in mammals [2]. The CPSase domain is located between an N-terminal GATase domain and the C-terminal part which encompass the dihydroorotase and aspartate transcarbamylase activities.

Ammonia-dependent CPSase (CPSase I) is involved in the urea cycle in ureolytic vertebrates; it is a monofunctional protein located in the mitochondrial matrix.

The CPSase domain is typically 120 Kd in size and has arisen from the duplication of an ancestral subdomain of about 500 amino acids. Each subdomain independently binds to ATP and it is suggested that the two homologous halves act separately, one to catalyze the phosphorylation of bicarbonate to carboxy phosphate and the other that of carbamate to carbamyl phosphate.

The CPSase subdomain is also present in a single copy in the biotin-dependent enzymes acetyl-CoA carboxylase (EC 6.4.1.2) (ACC), propionyl-CoA carboxylase (EC 6.4.1.3) (PCCase), pyruvate carboxylase (EC 6.4.1.1) (PC) and urea carboxylase (EC 6.3.4.6).

Two conserved regions which are probably important for binding ATP and/or catalytic activity have been selected as signatures for the subdomain.

-Consensus pattern: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG]

-Consensus pattern: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC]

[1] Simmer J.P., Kelly R.E., Rinker A.G. Jr., Scully J.L., Evans D.R.
J. Biol. Chem. 265:10395-10402(1990).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B.
BioEssays 15:157-164(1993).

87. CRAL_TRIO (CRAL/TRIO domain)

[1]

Medline: 98121119

Crystal structure of the Saccharomyces cerevisiae phosphatidyl-inositol-transfer protein.

Sha B, Phillips SE, Bankaitis VA, Luo M;

Nature 1998;391:506-510.

The original profile has been extended to include the carboxyl domain from the known structure of Sec14. Swiss:P10911 has not been included in the Pfam family because it does not appear to contain a complete structural domain.

Number of members: 39

88. CSD ('Cold-shock' DNA-binding domain)

[1]

Medline: 94255482

Crystal structure of CspA, the major cold shock protein of Escherichia coli.

Schindelin H, Jiang W, Inouye M, Heinemann U;

Proc Natl Acad Sci U S A 1994;91:5119-5123.

Number of members: 121

A conserved domain of about 70 amino acids has been found in prokaryotic and eukaryotic DNA-binding proteins [1,2,3,E1]. This domain, which is known as the 'cold-shock domain' (CSD) is present in the proteins listed below.

- Escherichia coli protein CS7.4 (gene cspA) which is induced in response to low temperature (cold-shock protein) and which binds to and stimulates the transcription of the CCAAT-containing promoters of the HN-S protein and of gyrA.
- Mammalian Y box binding protein 1 (YB1). A protein that binds to the CCAAT-containing Y box of mammalian HLA class II genes.
- Xenopus Y box binding proteins -1 and -2 (Y1 and Y2). Proteins that bind to the CCAAT-containing Y box of Xenopus hsp70 genes.
- Xenopus B box binding protein (YB3). YB3 binds the B box promoter element of genes transcribed by RNA polymerase III.
- Enhancer factor I subunit A (EFI-A) (dbpB). A protein that also bind to CCAAT-motif in various gene promoters.
- DbpA, a Human DNA-binding protein of unknown specificity.
- Bacillus subtilis cold-shock proteins cspB and cspC.
- Streptomyces clavuligerus protein SC 7.0.
- Escherichia coli proteins cspB, cspC, cspD, cspE and cspF.
- Unr, a mammalian gene encoded upstream of the N-ras gene. Unr contains nine repeats that are similar to the CSD domain. The function of Unr is not yet known but it could be a multivalent DNA-binding protein.

As a signature pattern for the CSD domain, its most conserved region which is located in its N-terminal section has been selected. It must be noted that the beginning of this region is highly similar [4] to the RNP-1 RNA-binding motif.

-Consensus pattern: [FY]-G-F-I-x(6,7)-[DER]-[LIVM]-F-x-H-x-[STKR]-x-[LIVMFY]

[1] Doniger J., Landsman D., Gonda M.A., Wistow G.
New Biol. 4:389-395(1992).
[2] Wistow G.
Nature 344:823-824(1990).
[3] Jones P.G., Inouye M.
Mol. Microbiol. 11:811-818(1994).
[4] Landsman D.
Nucleic Acids Res. 20:2861-2864(1992).

89. CTF_NFI (CTF/NF-I family)
Number of members: 45

Nuclear factor I (NF-I) or CCAAT box-binding transcription factor (CTF) [1,2] (also known as TGGCA-binding proteins) are a family of vertebrate nuclear proteins which recognize and bind, as dimers, the palindromic DNA sequence 5'-TGGCANNNTGCCA-3'. CTF/NF-I binding sites are present in viral and cellular promoters and in the origin of DNA replication of Adenovirus type 2.

The CTF/NF-I proteins were first identified as nuclear factor I, a collection of proteins that activate the replication of several Adenovirus serotypes (together with NF-II and NF-III) [3]. The family of proteins was also identified as the CTF transcription factors, before the NFI and CTF families were found to be identical [4]. The CTF/NF-I proteins are individually capable of activating transcription and DNA replication. The CTF/NF-I family name has also been dubbed as NFI, NF-I or NF1.

In a given species, there are a large number of different CTF/NF-I proteins. The multiplicity of CTF/NF-I is known to be generated both by alternative splicing and by the occurrence of four different genes. The known forms of NF-I genes have been classified as:

- The CTF-like factors subfamily (prototype form: CTF-1) [4]
- The NFI-X proteins.
- The NFI-A proteins.
- The NFI-B proteins.

So far, all CTF/NF-I family members appear to have similar transcription and replication activities.

CTF/NF-1 proteins contains 400 to 600 amino acids. The N-terminal 200 amino-acid sequence, almost perfectly conserved in all species and genes sequenced, mediates site-specific DNA recognition, protein dimerization and Adenovirus DNA replication. The C-terminal 100 amino acids contain the transcriptional activation domain. This activation domain is the target of gene expression regulatory pathways ellicited by growth factors and it interacts with basal transcription factors and with histone H3 [6].

A perfectly conserved, highly charged 12 residue peptide located in the N-terminal part of CTF/NF-I has been selected as a specific signature for this family of proteins.

-Consensus pattern: R-K-R-K-Y-F-K-K-H-E-K-R

[1] Mermod N., O'Neill E.A., Kelly T.J., Tjian R.
Cell 58:741-753(1989).
[2] Rupp R.A.W., Kruse U., Multhaup G., Goebel U., Beyreuther K., Sippel A.E.
Nucleic Acids Res. 18:2607-2616(1990).
[3] Nagata K., Guggenheimer R.A., Enomoto T., Lichy J.H., Hurwitz J.
Proc. Natl. Acad. Sci. U.S.A. 79:6438-6442(1982).
[4] Santoro C., Mermod N., Andrews P.C., Tjian R.
Nature 334:2118-2224(1988).
[5] Gil G., Smith J.R., Goldstein J.L., Slaughter C.A., Orth K., Brown M.S., Osborne T.F.
Proc. Natl. Acad. Sci. U.S.A 85:8963-8967(1988).

[6] Alevizopoulos A., Dusserre Y., Tsai-Pflugfelder M., von der Weid T., Wahli W., Mermod N.
Genes Dev. 9:3051-3066(1995).

90. Calsequestrin (Calsequestrin)

Number of members: 13

Calsequestrin is a moderate-affinity, high-capacity calcium-binding protein of cardiac and skeletal muscle [1], where it is located in the lumenal space of the sarcoplasmic reticulum terminal cisternae. Calsequestrin acts as a calcium buffer and plays an important role in the muscle excitation-contraction coupling. It is a highly acidic protein of about 400 amino acid residues that binds more than 40 moles of calcium per mole of protein. There are at least two different forms of calsequestrin: one which is expressed in cardiac muscles and another in skeletal muscles. Both forms have highly similar sequences.

Two signature sequences have been developed. The first corresponds to the N-terminus of the mature protein, the second is located just in front of the C-terminus of the protein which is composed of a highly acidic tail of variable length.

-Consensus pattern: [EQ]-[DE]-G-L-[DN]-F-P-x-Y-D-G-x-D-R-V

-Consensus pattern: [DE]-L-E-D-W-[LIVM]-E-D-V-L-x-G-x-[LIVM]-N-T-E-D-D-D

[1] Treves S., Vilsen B., Chiozzi P., Andersen J.P., Zorzato F.
Biochem. J. 283:767-772(1992).

91. Carboxyl_trans (Carboxyl transferase domain)

[1]

Medline: 93374821

Primary structure of the monomer of the 12S subunit of

transcarboxylase as deduced from DNA and characterization of the product expressed in Escherichia coli.

Thornton CG, Kumar GK, Haase FC, Phillips NF, Woo SB, Park VM, Magner WJ, Shenoy BC, Wood HG, Samols D;

J Bacteriol 1993;175:5301-5308.

[2]

Medline: 93358891

Molecular evolution of biotin-dependent carboxylases.

Toh H, Kondo H, Tanabe T;

Eur J Biochem 1993;215:687-696.

All of the members in this family are biotin dependent carboxylases. The carboxyl transferase domain carries out the following reaction; transcarboxylation from biotin to an acceptor molecule. There are two recognised types of carboxyl transferase. One of them uses acyl-CoA and the other uses 2-oxo acid as the acceptor molecule of carbon dioxide. All of the members in this family utilise acyl-CoA as the acceptor molecule.

Number of members: 47

92. Chal_stil_synt (Chalcone and stilbene synthases)

Number of members: 146

Chalcone synthases (CHS) (EC 2.3.1.74) and stilbene synthases (STS) (formerly known as resveratrol synthases) are related plant enzymes [1]. CHS is an important enzyme in flavanoid biosynthesis and STS a key enzyme in stilbene-type phyloalexin biosynthesis. Both enzymes catalyze the addition of three molecules of malonyl-CoA to a starter CoA ester (a typical example is 4-coumaroyl-CoA), producing either a chalcone (with CHS) or stilbene (with STS).

These enzymes are proteins of about 390 amino-acid residues. A conserved cysteine residue, located in the central section of these proteins, has been shown [2] to be essential for the catalytic activity of both enzymes and

probably represents the binding site for the 4-coumaryl-CoA group. The region around this active site residue is well conserved and can be used as a signature pattern.

In addition to the plant enzymes, this family also includes Bacillus subtilis bcsA.

-Consensus pattern: R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GA]-G-[GA]-[STAV]-x-[LIVMF]-[RA] [C is the active site residue]

[1] Schroeder J., Schroeder G.
Z. Naturforsch. 45C:1-8(1990).
[2] Lanz T., Tropf S., Marner F.-J., Schroeder J., Schroeder G.
J. Biol. Chem. 266:9971-9976(1991).

93. Chorismate_synt (Chorismate synthase)
Number of members: 19

Chorismate synthase (EC 4.6.1.4) catalyzes the last of the seven steps in the shikimate pathway which is used in prokaryotes, fungi and plants for the biosynthesis of aromatic amino acids. It catalyzes the 1,4-trans elimination of the phosphate group from 5-enolpyruvylshikimate-3-phosphate (EPSP) to form chorismate which can then be used in phenylalanine, tyrosine or tryptophan biosynthesis. Chorismate synthase requires the presence of a reduced flavin mononucleotide (FMNH2 or FADH2) for its activity.

Chorismate synthase from various sources shows [1,2] a high degree of sequence conservation. It is a protein of about 360 to 400 amino-acid residues. Three signature patterns have been developed from conserved regions rich in basic residues (mostly arginines). The first is in the N-terminal section, the second is central and the third is C-terminal.

-Consensus pattern: G-E-S-H-[GC]-x(2)-[LIVM]-[GTV]-x-[LIVM](2)-[DE]-G-x-[PV]

-Consensus pattern: [GE]-R-[SA](2)-[SAG]-R-[EV]-[ST]-x(2)-[RH]-V-x(2)-G

-Consensus pattern: R-[SH]-D-[PSV]-[CSAV]-x(4)-[GAI]-x-[IVGSP]-[LIVM]-x-E-[STAH]-[LIVM]

[1] Schaller A., Schmid J., Leibinger U., Amrhein N.
J. Biol. Chem. 266:21434-21438(1991).

[2] Jones D.G.L., Reusser U., Braus G.H.
Mol. Microbiol. 5:2143-2152(1991).

94. Clat_adaptor_s (Clathrin adaptor complex small chain)

Number of members: 21

Clathrin coated vesicles (CCV) mediate intracellular membrane traffic such as receptor mediated endocytosis. In addition to clathrin, the CCV are composed of a number of other components including oligomeric complexes which are known as adaptor or clathrin assembly proteins (AP) complexes [1]. The adaptor complexes are believed to interact with the cytoplasmic tails of membrane proteins, leading to their selection and concentration. In mammals two type of adaptor complexes are known: AP-1 which is associated with the Golgi complex and AP-2 which is associated with the plasma membrane. Both AP-1 and AP-2 are heterotetramers that consist of two large chains - the adaptins - (gamma and beta' in AP-1; alpha and beta in AP-2); a medium chain (AP47 in AP-1; AP50 in AP-2) and a small chain (AP19 in AP-1; AP17 in AP-2).

The small chains of AP-1 and AP-2 are evolutionary related proteins of about 18 Kd. Homologs of AP17 and AP19 have also been found in yeast (genes APS1/ YAP19 and APS2/YAP17) [2,3,4]. AP17 and AP19 are also related to the zeta-chain [5] of coatomer (zeta-cop), a cytosolic protein complex that reversibly associates with Golgi membranes to form vesicles that mediate biosynthetic protein transport from the endoplasmic reticulum, via the Golgi up to the trans Golgi network.

A conserved region in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM](2)-Y-[KR]-x(4)-L-Y-F

[1] Pearse B.M., Robinson M.S.
Annu. Rev. Cell Biol. 6:151-171(1990).
[2] Kirchhausen T., Davis A.C., Frucht S., O'Brine Greco B., Payne G.S., Tubb B.
J. Biol. Chem. 266:11153-11157(1991).
[3] Nakai M., Takada T., Endo T.
Biochim. Biophys. Acta 1174:282-284(1993).
[4] Phan H.L., Finlay J.A., Chu D.S., Tan P.K., Kirchhausen T., Payne G.S.
EMBO J. 13:1706-1717(1994).
[5] Kuge O., Hara-Kuge S., Orci L., Ravazzola M., Amherdt M., Tanigawa G., Wieland F.T., Rothman J.E.
J. Cell Biol. 123:1727-1734(1993).

95. Clathrin_lg_ch (Clathrin light chain.)
Number of members: 8

Clathrin [1,2] is the major coat-forming protein that encloses vesicles such as coated pits and forms cell surface patches involved in membrane traffic within eukaryotic cells. The clathrin coats (called triskelions) are composed of three heavy chains (180 Kd) and three light chains (23 to 27 Kd).

The clathrin light chains [3], which may help to properly orient the assembly and disassembly of the clathrin coats, bind non-covalently to the heavy chain, they also bind calcium and interact with the hsc70 uncoating ATPase.

- In higher eukaryotes two genes code for distinct but related light chains: LC(a) and LC(b). Each of the two genes can yield, by tissue-specific alternative splicing, two separate forms which differ by the insertion of a

sequence of respectively thirty or eighteen residues. There is, in the N-terminal part of the clathrin light chains a domain of twenty one amino acid residues which is perfectly conserved in LC(a) and LC(b).

- In yeast there is a single light chain (gene CLC1) whose sequence is only distantly related to that of higher eukaryotes.

Two signature patterns have been developed for clathrin light chains. The first pattern is a heptapeptide from the center of the conserved N-terminal region of eukaryotic light chains; the second pattern is derived from a positively charged region located in the C-terminal extremity of all known clathrin light chains.

-Consensus pattern: F-L-A-Q-Q-E-S

[1] Keen J.H.
Annu. Rev. Biochem. 59:415-438(1990).
[2] Brodsky F.M.
Science 242:1396-1402(1988).
[3] Brodsky F.M., Hill B.L., Acton S.L., Naethke I., Wong D.H., Ponnambalam S., Parham P.
Trends Biochem. Sci. 16:208-213(1991).

96. (Clathrin repeat) 7-fold repeat in Clathrin and VPS

Each repeat is about 140 amino acids long. The repeats occur in the arm region of the Clathrin heavy chain.

Number of members: 79

[1]
Medline: 92191269
Folding and trimerization of clathrin subunits at the triskelion hub.
Nathke IS, Heuser J, Lupas A, Stock J, Turck CW, Brodsky FM;
Cell 1992;68:899-910. [2]
Medline: 88097376

Clathrin heavy chain: molecular cloning and complete primary structure.

Kirchhausen T, Harrison SC, Chow EP, Mattaliano RJ, Ramachandran KL, Smart J, Brosius J;

Proc Natl Acad Sci U S A 1987;84:8805-8809.

97. Collagen (Collagen triple helix repeat (20 copies))

[1] Medline: 94059583

New members of the collagen superfamily

Mayne R, Brewton RG;

Curr Opin Cell Biol 1993;5:883-890.

Scurvy is associated with collagens.

Members of this family belong to the collagen superfamily [1]. Collagens are generally extracellular structural proteins involved in formation of connective tissue structure. The alignment contains 20 copies of the G-X-Y repeat that forms a triple helix. The first position of the repeat is glycine, the second and third positions can be any residue but are frequently proline and hydroxyproline. Collagens are post translationally modified by proline hydoxylase to form the hydroxyproline residues. Defective hydroxylation is the cause of scurvy. Some members of the collagen superfamily are not involved in connective tissue structure but share the same triple helical structure.

Number of members: 2125

98. Coprogen_oxidas (Coproporphyrinogen III oxidase)

Number of members: 12

Coproporphyrinogen III oxidase (EC 1.3.3.3) (coproporphyrinogenase) [1,2] catalyzes the oxidative decarboxylation of coproporphyrinogen III into protoporphyrinogen IX, a common step in the pathway for the biosynthesis of

porphyrins such as heme, chlorophyll or cobalamin.

Coproporphyrinogen III oxidase is an enzyme that requires iron for its activity. A cysteine seems to be important for the catalytic mechanism [3]. Sequences from a variety of eukaryotic and prokaryotic sources show that this enzyme has been evolutionarily conserved. A highly conserved region in the central part of the sequence has been selected as a signature pattern. This region contains the only conserved cysteine and is rich in charged amino acids.

-Consensus pattern: K-x-W-C-x(2)-[FYH](3)-[LIVM]-x-H-R-x-E-x-R-G-[LIVM]-G-G-[LIVM]-F-F-D

[1] Xu K., Elliott T.
J. Bacteriol. 175:4990-4999(1993).
[2] Kohno H., Furukawa T., Yoshinaga T., Tokunaga R., Taketani S.
J. Biol. Chem. 268:21359-21363(1993).
[3] Camadro J.M., Chambon H., Jolles J., Labbe P.
Eur. J. Biochem. 156:579-587(1986).
[4] Xu K., Elliott T.
J. Bacteriol. 176:3196-3203(1994).

99. Corona_nucleoca (Coronavirus nucleocapsid protein)

[1]
Medline: 98087828
Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal.
Molenkamp R, Spaan WJ;
Virology 1997;239:78-86.
Number of members: 44

100. Cu-oxidase (Multicopper oxidase)

[1]

Medline: 90126844

The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin. Modelling and structural relationships.

Messerschmidt A, Huber R;

Eur J Biochem 1990;187:341-352.

Number of members: 150

Multicopper oxidases [1,2] are enzymes that possess three spectroscopically different copper centers. These centers are called: type 1 (or blue), type 2 (or normal) and type 3 (or coupled binuclear). The enzymes that belong to this family are:

- Laccase (EC 1.10.3.2) (urishiol oxidase), an enzyme found in fungi and plants, which oxidizes many different types of phenols and diamines.
- Ascorbate oxidase (EC 1.10.3.3), a higher plant enzyme.
- Ceruloplasmin (EC 1.16.3.1) (ferroxidase), a protein found in the serum of mammals and birds, which oxidizes a great variety of inorganic and organic substances. Structurally ceruloplasmin exhibits internal sequence homology, and seem to have evolved from the triplication of a copper-binding domain similar to that found in laccase and ascorbate oxidase.

In addition to the above enzymes there are a number of proteins which, on the basis of sequence similarities, can be said to belong to this family. These proteins are:

- Copper resistance protein A (copA) from a plasmid in Pseudomonas syringae. This protein seems to be involved in the resistance of the microbial host to copper.
- Blood coagulation factor V (Fa V).
- Blood coagulation factor VIII (Fa VIII) [E1].
- Yeast FET3 [3], which is required for ferrous iron uptake.
- Yeast hypothetical protein YFL041w and SpAC1F7.08, the fission yeast

homolog.

Factors V and VIII act as cofactors in blood coagulation and are structurally similar [4]. Their sequence consists of a triplicated A domain, a B domain and a duplicated C domain; in the following order: A-A-B-A-C-C. The A-type domain is related to the multicopper oxidases.

Two signature patterns have been developed for these proteins. Both patterns are derived from the same region, which in ascorbate oxidase, laccase, in the third domain of ceruloplasmin, and in copA, contains five residues that are known to be involved in the binding of copper centers. The first pattern does not make any assumption on the presence of copper-binding residues and thus can detect domains that have lost the ability to bind copper (such as those in Fa V and Fa VIII), while the second pattern is specific to copper-binding domains.

-Consensus pattern: G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x(8)-G-[LM]-x(3)-[LIVMFYW]
-Consensus pattern: H-C-H-x(3)-H-x(3)-[AG]-[LM]
[The first two H's are copper type 3 binding residues]
[The C, the 3rd H, and L or M are copper type 1 ligands]

101. Cullin (Cullin family)
Number of members: 24

The following proteins are collectively termed cullins [1]:

- Caenorhabditis elegans cul-1 (or lin-19), a protein required for developmentally programmed transitions from the G1 phase of the cell cycle to the G0 phase or the apoptotic pathway.
- Caenorhabditis elegans cul-2, cul-3, cul-4 (F45E12.3), cul-5 (ZK856.1) and cul-6 (K08E7.7).
- Mammalian CUL1, CUL2, CUL3, CUL4A and CUL4B.
- Mammalian vasopressin-activated calcium-mobilizing receptor (VACM-1), a

kidney-specific protein thought to form a cell surface receptor [2] but which does not have any structural hallmarks of a receptor.

- Drosophila lin19.
- Yeast CDC53 [3], which acts in concert with CDC4 and UBC3 (CDC34) to control the G1-to-S phase transition.
- Yeast hypothetical protein YGR003w.
- Fission yeast hypothetical protein SpAC24H6.03.

The cullins are hydrophilic proteins of 740 to 815 amino acids. The C-terminal extremity is the most conserved part of these proteins. A signature pattern has been developed from that region.

-Consensus pattern: [LIV]-K-x(2)-[LIV]-x(2)-L-I-[DEQ]-[KRHNQ]-x-Y-[LIVM]-x-R-x(6,7)-[FY]-x-Y-x-[SA]>

[1] Kipreos E.T., Lander L.E., Wing J.P., He W.W., Hedgecock E.M.
Cell 85:829-839(1996).
[2] Burnatowska-Hledin M.A., Spielman W.S., Smith W.L., Shi P., Meyer J.M., Dewitt D.L.
Am. J. Physiol. 268:f1198-F1210(1995).
[3] Mathias N., Johnson S.L., Winey M., Adams A.E., Goetsch L., Pringle J.R., Byers B., Goebl M.G.
Mol. Cell. Biol. 16:6634-6643(1996).

102. (Cu_amine_oxid)

Copper amine oxidase signatures

Amine oxidases (AO) [1] are enzymes that catalyze the oxidation of a wide range of biogenic amines including many neurotransmitters, histamine and xenobiotic amines. There are two classes of amine oxidases: flavin-containing (EC 1.4.3.4) and copper-containing (EC 1.4.3.6).

Copper-containing AO is found in bacteria, fungi, plants and animals, it is an homodimeric enzyme that binds one copper ion per subunit as well as a 2,4,5- trihydroxyphenylalanine quinone (or topaquinone) (TPQ) cofactor. This cofactor is derived from a tyrosine residue.

Two signature patterns were derived for copper AO, the first one contains the tyrosine which give rises to the TPQ cofactor while the second one contains one of the three histidines that bind the copper atom [2].

Consensus pattern[LIVM]-[LIVMA]-[LIVMF]-x(4)-[ST]-x(2)-N-Y-[DE]-[YN] [The first Y gives rises to TPQ] Sequences known to belong to this class detected by the patternALL.

Consensus patternT-x-[GS]-x(2)-H-[LIVMF]-x(3)-E-[DE]-x-P [H is a copper ligand] Sequences known to belong to this class detected by the pattern ALL, except for lentil AO.

[1] Knowles P.F., Dooley D.M. (In) Metal ions in biological systems; Sigel H., Sigel A., Eds., 30:361- 403, Marcel Dekker, New-York, (1993).
[2] Parsons M.R., Convery M.A., Wilmot C.M., Yadav K.D.S., Blakeley V., Corner A.S., Phillips S.E.V., McPherson M.J., Knowles P.F. Structure 3:1171-1184(1995).

103. Cys-protease (Cysteine protease)
Number of members: 358

Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences).

- Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2].
- Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2].
- Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium-activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain.

- Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3].
- Human cathepsin O [4].
- Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide).
- Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A.
- House-dust mites allergens DerP1 and EurM1.
- Cathepsin B-like proteinases from the worms Caenorhabditis elegans (genes gcp-1, cpr-3, cpr-4, cpr-5 and cpr-6), Schistosoma mansoni (antigen SM31) and Japonica (antigen SJ31), Haemonchus contortus (genes AC-1 and AC-2), and Ostertagia ostertagi (CP-1 and CP-3).
- Slime mold cysteine proteinases CP1 and CP2.
- Cruzipain from Trypanosoma cruzi and brucei.
- Throphozoite cysteine proteinase (TCP) from various Plasmodium species.
- Proteases from Leishmania mexicana, Theileria annulata and Theileria parva.
- Baculoviruses cathepsin-like enzyme (v-cath).
- Drosophila small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain.
- Yeast thiol protease BLH1/YCP1/LAP3.
- Caenorhabditis elegans hypothetical protein C06G4.2, a calpain-like protein.

Two bacterial peptidases are also part of this family:

- Aminopeptidase C from Lactococcus lactis (gene pepC) [5].
- Thiol protease tpr from Porphyromonas gingivalis.

Three other proteins are structurally related to this family, but may have

lost their proteolytic activity.

- Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine.
- Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see <PDOC00382>).
- Plasmodium falciparum serine-repeat protein (SERA), the major blood stage antigen. This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine.

The sequences around the three active site residues are well conserved and can be used as signature patterns.

-Consensus pattern: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]
-Consensus pattern: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH] [H is the active site residue]

-Consensus pattern: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue]

[1] Dufour E. Biochimie 70:1335-1342(1988).
[2] Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).
[3] Shi G.-P., Chapman H.A., Bhairi S.M., Deleeuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).
[4] Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).
[5] Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).
[6] Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).
[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

104. Cys_Met_Meta_PP (Cys/Met metabolism PLP-dependent enzyme)

[1] Medline: 96428687

Crystal structure of the pyridoxal-5'-phosphate dependent cystathionine beta-lyase from Escherichia coli at 1.83 A.

Clausen T, Huber R, Laber B, Pohlenz HD, Messerschmidt A;

J Mol Biol 1996;262:202-224.

[1] Medline: 99059720

Crystal structure of Escherichia coli cystathionine gamma-synthase at 1.5 A resolution.

Clausen T, Huber R, Prade L, Wahl MC, Messerschmidt A;

EMBO J 1998;17:6827-6838.

Database Reference: SCOP; 1cs1; fa; [SCOP-USA][CATH-PDBSUM]

This family includes enzymes involved in cysteine and methionine metabolism. The following are members:

Cystathionine gamma-lyase,

Cystathionine gamma-synthase,

Cystathionine beta-lyase,

Methionine gamma-lyase,

OAH/OAS sulfhydrylase,

O-succinylhomoserine sulphhydrylase

All of these members participate is slightly different reactions.

All these enzymes use PLP (pyridoxal-5'-phosphate) as a cofactor.

Number of members: 52

A number of pyridoxal-dependent enzymes involved in the metabolism of cysteine, homocysteine and methionine have been shown [1,2] to be evolutionary related. These are:

- Cystathionine gamma-lyase (EC 4.4.1.1) (gamma-cystathionase), which catalyzes the transformation of cystathionine into cysteine, oxobutanoate and ammonia. This is the final reaction in the transulfuration pathway that leads from methionine to cysteine in eukaryotes.
- Cystathionine gamma-synthase (EC 4.2.99.9), which catalyzes the conversion

of cysteine and succinyl-homoserine into cystathionine and succinate: the first step in the biosynthesis of methionine from cysteine in bacteria (gene metB).

- Cystathionine beta-lyase (EC 4.4.1.8) (beta-cystathionase), which catalyzes the conversion of cystathionine into homocysteine, pyruvate and ammonia: the second step in the biosynthesis of methionine from cysteine in bacteria (gene metC).
- Methionine gamma-lyase (EC 4.4.1.11) (L-methioninase) which catalyzes the transformation of methionine into methanethiol, oxobutanoate and ammonia.
- OAH/OAS sulfhydrylase, which catalyzes the conversion of acetylhomoserine into homocysteine and that of acetylserine into cysteine (gene MET17 or MET25 in yeast).
- O-succinylhomoserine sulfhydrylase (EC 4.2.99.-).
- Yeast hypothetical protein YGL184c.
- Yeast hypothetical protein YHR112c.

These enzymes are proteins of about 400 amino-acid residues. The pyridoxal-P group is attached to a lysine residue located in the central section of these enzymes; the sequence around this residue is highly conserved and can be used as a signature pattern to detect this class of enzymes.

-Consensus pattern: [DQ]-[LIVMF]-x(3)-[STAGC]-[STAGCI]-T-K-[FYWQ]-[LIVMF]-x-G-[HQ]-[SGNH] [K is the pyridoxal-P attachment site]

[1] Ono B.I., Tanaka K., Naito K., Heike C., Shinoda S., Yamamoto S., Ohmori S., Oshima T., Toh-E A.
J. Bacteriol. 174:3339-3347(1992).

[2] Barton A.B., Kaback D.B., Clark M.W., Keng T., Ouellette B.F.F., Storms R.K., Zeng B., Zhong W.W., Fortin N., Delaney S., Bussey H.
Yeast 9:363-369(1993).

105. Cyt_reductase

FAD/NAD-binding Cytochrome reductase

Number of members: 60

[1] Medline: 95111952

Crystal structure of the FAD-containing fragment of corn nitrate reductase at 2.5 A resolution: relationship to other flavoprotein reductases.

Lu G, Campbell WH, Schneider G, Lindqvist Y;

Structure 1994;2:809-821.

[2] Medline: 92084635

The sequence of squash NADH:nitrate reductase and its relationship to the sequences of other flavoprotein oxidoreductases. A family of flavoprotein pyridine nucleotide cytochrome reductases.

Hyde GE, Crawford NM, Campbell WH;

J Biol Chem 1991;266:23542-23547.

106. Cytidylyltrans

Phosphatidate cytidylyltransferase

Number of members: 21

Phosphatidate cytidylyltransferase (EC 2.7.7.41) [1,2,3] (also known as CDP-diacylglycerol synthase) (CDS) is the enzyme that catalyzes the synthesis of CDP-diacylglycerol from CTP and phosphatidate (PA). CDP-diacylglycerol is an important branch point intermediate in both prokaryotic and eukaryotic organisms. CDS is a membrane-bound enzyme. A conserved region located in the C-terminal part has been selected as a signature pattern.

-Consensus pattern: S-x-[LIVMF]-K-R-x(4)-K-D-x-[GSA]-x(2)-[LI]-[PG]-x-H-G-G-[LIVM]-x-D-R-[LIVMF]-D

[1] Sparrow C.P., Raetz C.R.H.

J. Biol. Chem. 260:12084-12091(1985).

[2] Shen H., Heacock P.N., Clancey C.J., Dowhan W.

J. Biol. Chem. 271:789-795(1996).

[3] Saito S., Goto K., Tonosaki A., Kondo H.
J. Biol. Chem. 272:9503-9509(1997).

107. (Cytidylyltransf) Cytidylyltransferase. This family includes: Cholinephosphate cytidylyltransferase. Glycerol-3-phosphate cytidylyltransferase.
Number of members: 64

[1] Medline: 10208837 CTP:Phosphocholine Cytidylyltransferase: Insights into Regulatory Mechanisms and Novel Functions. Clement JM, Kent C; Biochem Biophys Res Commun 1999;257:643-650.

108. (cNMP binding) Cyclic nucleotide-binding domain signatures and profile
Proteins that bind cyclic nucleotides (cAMP or cGMP) share a structural domainof about 120 residues [1-3]. The best studied of these proteins is theprokaryotic catabolite gene activator (also known as the cAMP receptorprotein) (gene crp) where such a domain is known to be composed of threealpha-helices and a distinctive eight-stranded, antiparallel beta-barrelstructure. Such a domain is known to exist in the following proteins: - Prokaryotic catabolite gene activator protein (CAP). - cAMP- and cGMP-dependent protein kinases (cAPK and cGPK). Both types of kinases contains two tandem copies of the cyclic nucleotide-binding domain. The cAPK's are composed of two different subunits: a catalytic chain and a regulatory chain which contains both copies of the domain. The cGPK's are single chain enzymes that include the two copies of the domain in their N- terminal section. The nucleotide specificity of cAPK and cGPK is due to an amino acid in the conserved region of beta-barrel 7: a threonine that is invariant in cGPK is an alanine in most cAPK. - Vertebrate cyclic nucleotide-gated ion-channels. Two such cations channels have been fully characterized. One is found in rod cells where it plays a role in visual signal transduction. It specifically binds to cGMP leading to an opening of the channel and thereby causing a depolarization of rod photoreceptors. In olfactory epithelium a similar, cAMP-binding, channel plays a role in odorant signal transduction. There are six invariant amino acids in this domain, three of which are glycine residues that are thought to be essential for maintenance of the of the beta-barrel. Two signature patterns for this domain have been developed. The first pattern is located within beta-barrels 2 and 3 and contains the first two

conserved Gly. The second pattern is located within beta-barrels 6 and 7 and contains the third conserved Gly as well as the three other invariant residues.-

First consensus pattern: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G

Second consensus pattern: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x- [STACV]-

[1] Weber I.T., Shabb J.B., Corbin J.D. Biochemistry 28:6122-6127(1989).

[2] Kaupp U.B. Trends Neurosci. 14:150-157(1991).

[3] Shabb J.B., Corbin J.D. J. Biol. Chem. 267:5723-5726(1992).

109. (cadherin)

Cadherins extracellular repeated domain signature

Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known:

- Epithelial (E-cadherin) (also known as uvomorulin or L-CAM) (CDH1).
- Neural (N-cadherin) (CDH2).
- Placental (P-cadherin) (CDH3).
- Retinal (R-cadherin) (CDH4).
- Vascular endothelial (VE-cadherin) (CDH5).
- Kidney (K-cadherin) (CDH6).
- Cadherin-8 (CDH8).
- Osteoblast (OB-cadherin) (CDH11).
- Brain (BR-cadherin) (CDH12).
- T-cadherin (truncated cadherin) (CDH13).
- Muscle (M-cadherin) (CDH14).
- Liver-intestine (LI-cadherin).
- EP-cadherin.

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub- divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

The Drosophila fat protein [3] is a huge protein of over 5000 amino acids that contains 34 cadherin-like repeats in its extracellular domain.

The signature pattern that was developed for the repeated domain is located in it the C-terminal extremity which is its best conserved region. The pattern includes two conserved aspartic acid residues as well as two asparagines; these residues could be implicated in the binding of calcium.

Consensus pattern[LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P Sequences known to belong to this class detected by the pattern ALL. Note this pattern is found in the first, second, and fourth copies of the repeated domain. In the third copy there is a deletion of one residue after the second conserved Asp.

[1] Takeichi M. Annu. Rev. Biochem. 59:237-252(1990).
[2] Takeichi M. Trends Genet. 3:213-217(1987).
[3] Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J., Goodman C.S. Cell 67:853-868(1991).

110. Calreticulin family signatures

Calreticulin [1] (also known as calregulin, CRP55 or HACBP) is a high-capacitycalcium-binding protein which is present in most tissues and located at the periphery of the endoplasmic (ER) and the sarcoplamic reticulum (SR)membranes. It probably plays a role in the storage of calcium in the lumen ofthe ER and SR and it may well have other important functions. Structurally, calreticulin is a protein of about 400 amino acid residues consisting of three domains: a) An N-terminal, probably globular, domain of about 180 amino acid residues (N-domain); b) A central domain of about 70 residues (P-domain) which contains three repeats of an acidic 17 amino acid motif. This region binds calcium with a low-capacity, but a high-affinity; c) A C-terminal domain rich in acidic residues and in lysine (C-domain). This region binds calcium with a high-capacity but a low-affinity. Calreticulin is evolutionary related to the following proteins: - Onchocerca volvulus antigen RAL-1. RAL-1 is highly similar to calreticulin, but possesses a C-terminal domain rich in lysine and arginine and lacks acidic residues and is therefore not expected to bind calcium in that region. - Calnexin [2]. A calcium-binding protein that interacts with newly synthesized glycoproteins in the endoplasmic reticulum. It seems to play a major role in the quality control apparatus of the ER by the retention of incorrectly folded proteins. - Calmegin [3] (or calnexin-T), a testis-specific calcium-binding protein highly similar to calnexin. Three signature patterns have been developed for this family of proteins. The first two patterns are based on conserved regions in the N-domain; the third pattern corresponds to positions 4 to 16 of the repeated motif in the P-domain.

Consensus pattern: [KRHN]-x-[DEQN]-[DEQNK]-x(3)-C-G-G-[AG]-[FY]-[LIVM]-[KN]-[LIVMFY](2)-

Consensus pattern: [LIVM](2)-F-G-P-D-x-C-[AG]-

Consensus pattern: [IV]-x-D-x-[DENST]-x(2)-K-P-[DEH]-D-W-[DEN]-

[1] Michalak M., Milner R.E., Burns K., Opas M. Biochem. J. 285:681-692(1992).

[2] Bergeron J.J.M., Brenner M.B., Thomas D.Y., Williams D.B. Trends Biochem. Sci. 19:124-128(1994).

[3] Watanabe D., Yamada K., Nishina Y., Tajima Y., Koshimizu U., Nagata A., Nishimune Y. J. Biol. Chem. 269:7744-7749(1994).

111. Eukaryotic-type carbonic anhydrases signature (carb_anhydrase)

Carbonic anhydrases (EC 4.2.1.1) (CA) [1,2,3,4] are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide. Eight enzymatic and evolutionary related forms of carbonic anhydrase are currently known to exist in vertebrates: three cytosolic isozymes (CA-I, CA-II and CA-III); two membrane-bound forms (CA-IV and CA-VII); a mitochondrial form (CA-V); a secreted salivary form (CA-VI); and a yet uncharacterized isozyme [5].In the alga Chlamydomonas reinhardtii, two CA isozymes have been sequenced[6]. They are periplasmic glycoproteins evolutionary related to vertebrate CAs. Some bacteria, such as Neisseria gonorrhoeae [7] also have a eukaryotic-type CA.CAs contain a single zinc atom bound to three conserved histidine residues. As a signature for CAs, a pattern has been developed which includes one of these zinc-binding histidines. Protein D8 from Vaccinia and other poxviruses is related to CAs but has lost two of the zinc-binding histidines as well as many otherwise conserved residues. This is also true of the N-terminal extracellular domain of some receptor-type tyrosine-protein phosphatases (see <PDOC00323>).

Consensus pattern: S-E-[HN]-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVMGA]-H-[LIVMFA](2) [The second H is a zinc ligand]-

Note: most prokaryotic CA's as well as plant chloroplast CA's belong to another, evolutionary distinct family of proteins (see <PDOC00586

[1] Deutsch H.F. Int. J. Biochem. 19:101-113(1987).
[2] Fernley R.T. Trends Biochem. Sci. 13:356-359(1988).
[3] Tashian R.E. BioEssays 10:186-192(1989).
[4] Edwards Y. Biochem. Soc. Trans. 18:171-175(1990).
[5] Skaggs L.A., Bergenhem N.C.H., Venta P.J., Tashian R.E. Gene 126:291-292(1993).
[6] Fujiwara S., Fukuzawa H., Tachiki A., Miyachi S. Proc. Natl. Acad. Sci. U.S.A. 87:9779-9783(1990).
[7] Huang S., Xue Y., Sauer-Eriksson E., Chirica L., Lindskog S.,Jonsson B.H. 2.3.CO;2-"J. Mol. Biol. 283:301-310(1998).

112. Caseins alpha/beta signature

Caseins [1] are the major protein constituent of milk. Caseins can be classified into two families; the first consists of the kappa-caseins, and the second groups the alpha-s1, alpha-s2, and beta-caseins. The alpha/beta caseins are a rapidly diverging family of proteins. However two regions are conserved: a cluster of phosphorylated serine residues and the signal

sequence. The signature pattern has been developed for this family of proteins based upon the last eight residues of the signal sequence.

Consensus pattern: C-L-[LV]-A-x-A-[LVF]-A –

[1] Holt C., Sawyer L. Protein Eng. 2:251-259(1988).

113. Catalase signatures

Catalase (EC 1.11.1.6) [1,2,3] is an enzyme, present in all aerobic cells,that decomposes hydrogen peroxide to molecular oxygen and water. Its main function is to protect cells from the toxic effects of hydrogen peroxide. In eukaryotic organisms and in some prokaryotes catalase is a molecule composed of four identical subunits. Each of the subunits binds one protoheme IX group. A conserved tyrosine serves as the heme proximal side ligand. The region around this residue has been used as a first signature pattern; it also includes a conserved arginine that participates in heme-binding. A conserved histidine has been shown to be important for the catalytic mechanism of the enzyme. The region around this residue has been selected as a second signature pattern.-

Consensus pattern: R-[LIVMFSTAN]-F-[GASTNP]-Y-x-D-[AST]-[QEH] [Y is the proximal heme-binding ligand]

Consensus pattern: [IF]-x-[RH]-x(4)-[EQ]-R-x(2)-H-x(2)-[GAS]-[GASTF]-[GAST] [H is an active site residue]

Note: some prokaryotic catalases belong to the peroxidase family (see <PDOC00394>).

[1] Murthy M.R.N., Reid T.J. III, Sicignano A., Tanaka N., Rossmann M.G. J. Mol. Biol. 152:465-499(1981).

[2] Melik-Adamyan W.R., Barynin V.V., Vagin A.A., Borisov V.V., Vainshtein B.K., Fita I., Murthy M.R.N., Rossmann M.G. J. Mol. Biol. 188:63-72(1986).

[3] von Ossowski I., Hausner G., Loewen P.C. J. Mol. Evol. 37:71-76(1993).

114. (chitin binding) Chitin recognition or binding domain signature

A conserved domain of 43 amino acids is found in several plant and fungal proteins that have a common binding specificity for oligosaccharides of N-acetylglucosamine [1]. This domain may be involved in the recognition or binding of chitin subunits. It has been found in the

proteins listed below. - A number of non-leguminous plant lectins. The best characterized of these lectins are the three highly homologous wheat germ agglutinins (WGA-1, 2 and 3). WGA is an N-acetylglucosamine/N-acetylneuraminic acid binding lectin which structurally consists of a fourfold repetition of the 43 amino acid domain. The same type of structure is found in a barley root-specific lectin as well as a rice lectin. - Plants endochitinases (EC 3.2.1.14) from class IA (see <PDOC00620>). Endochitinases are enzymes that catalyze the hydrolysis of the beta-1,4 linkages of N-acetyl glucosamine polymers of chitin. Plant chitinases function as a defense against chitin containing fungal pathogens. Class IA chitinases generally contain one copy of the chitin-binding domain at their N-terminal extremity. An exception is agglutinin/chitinase [2] from the stinging nettle Urtica dioica which contains two copies of the domain. - Hevein [5], a wound-induced protein found in the latex of rubber trees. - Win1 and win2, two wound-induced proteins from potato. - Kluyveromyces lactis killer toxin alpha subunit [3]. The toxin encoded by the linear plasmid pGKL1 is composed of three subunits: alpha, beta, and gamma. The gamma subunit harbors toxin activity and inhibits growth of sensitive yeast strains in the G1 phase of the cell cycle; the alpha subunit, which is proteolytically processed from a larger precursor that also contains the beta subunit, is a chitinase (see <PDOC00839>). In chitinases, as well as in the potato wound-induced proteins, the 43-residuedomain directly follows the signal sequence and is therefore at the N-terminal of the mature protein; in the killer toxin alpha subunit it is located in the central section of the protein. The domain contains eight conserved cysteine residues which have all been shown, in WGA, to be involved in disulfide bonds. The topological arrangement of the four disulfide bonds is shown in the following figure: +-------------+ +----|------+ | | | | | xxCgxxxxxxxCxxxxCCsxxgxCgxxxxxCxxxCxxxxC | ******|************ | | | | +----+ +--------------+'C': conserved cysteine involved in a disulfide bond.'*': position of the pattern.

-Consensus pattern: C-x(4,5)-C-C-S-x(2)-G-x-C-G-x(4)-[FYW]-C [The five C's are involved in disulfide bonds]

[1] Wright H.T., Sandrasegaram G., Wright C.S. J. Mol. Evol. 33:283-294(1991).
[2] Lerner D.R., Raikhel N.V. J. Biol. Chem. 267:11085-11091(1992).
[3] Butler A.R., O'Donnel R.W., Martin V.J., Gooday G.W., Stark M.J.R. Eur. J. Biochem. 199:483-488(1991).

115. (Chitinase 1) Chitinases family 19 signatures

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of thebeta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the viewpoint of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19(also known as classes IA or I and IB or II) are enzymes from plants that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (seethe relevant entry <PDOC00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues. Two highly conserved regions have been selected as signature patterns, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond.

Consensus pattern: C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F- [GSA]
Consensus pattern: [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM]

[1] Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992).
[2] Henrissat B. Biochem. J. 280:309-316(1991).

116. chloroa_b-bind

Chlorophyll A-B binding proteins. Number of members: 211

117. chromo

The 'chromo' (CHRromatin Organization MOdifier) domain [1 to 4] is a conserved region of about 60 amino acids which was originally found in Drosophila modifiers of variegation, which are proteins that modify the structure of chromatin to the condensed morphology of heterochromatin, a cytologically visible condition where gene expression is repressed. In protein Polycomb, the chromo domain has been shown to be important for chromatin targeting. Proteins that contains a chromo domain seem to fall into three classes:

a) Proteins which have a N-terminal chromo domain followed by a region which is related to but distinct from the chromo domain and which has been termed [3] the 'chromo shadow' domain.

b) Proteins with a single chromo domain.

c) Proteins with paired tandem chromo domains.

Currently, this domain has been found in the following proteins:

Class A.

- Drosophila heterochromatin protein Su(var)205 (HP1).
- Human heterochromatin protein HP1 alpha.
- Mammalian modifier 1 and modifier 2.
- Fission yeast swi6, a protein involved in the repression of the silent mating-type loci mat2 and mat3.

Class B.

- Drosophila protein Polycomb (Pc).
- Mammalian modifier 3, a homolog of Pc.
- Drosophila protein Su(var)3-9, a suppressor of position-effect variegation.
- Human Mi-2 autoantigen, characterisitic of dermatomyosis.
- Fungal retrotranposon polyproteins: 'skippy' from Fusarium oxysporum, 'grasshopper' and 'MAGGY' from Magnaporthe grisea and CfT-1 from Cladosporium fulvum.
- Fission yeast hypothetical protein SpAC18G6.02c.
- Caenorhabditis elegans hypothetical protein C29H12.5
- Caenorhabditis elegans hypothetical protein ZK1236.2.
- Caenorhabditis elegans hypothetical protein T09A5.8.

Class C.

- Mammalian DNA-binding/helicase proteins CHD-1 to CHD-4.
- Yeast protein CHD1.

The signature pattern for this domain corresponds to its best conserved section, which is located in its central part.

-Consensus pattern: [FYL]-x-[LIVMC]-[KR]-W-x-[GDNR]-[FYWLME]-x(5,6)-[ST]-W-[ESV]-[PSTDEN]-x(2,3)-[LIVMC]

[1] Paro R. Trends Genet. 6:416-421(1990).

[2] Singh P.B., Miller J.R., Pearce J., Kothary R., Burton R.D., Paro R., James T.C., Gaunt S.J. Nucleic Acids Res. 19:789-794(1991).

[3] Aasland R., Stewart A.F. Nucleic Acids Res. 23:3168-3173(1995).

[4] Koonin E.V., Zhou S., Lucchesis J.C. Nucleic Acids Res. 23:4229-4233(1995).

118. citrate_synt

Citrate synthase (EC 4.1.3.7) (CS) is the tricarboxylic acid cycle enzyme that catalyzes the synthesis of citrate from oxaloacetate and acetyl-CoA in an aldol condensation. CS can directly form a carbon-carbon bond in the absence of metal ion cofactors.

In prokaryotes, citrate synthase is composed of six identical subunits. In eukaryotes, there are two isozymes of citrate synthase: one is found in the mitochondrial matrix, the second is cytoplasmic. Both seem to be dimers of identical chains.

There are a number of regions of sequence similarity between prokaryotic and eukaryotic citrate synthases. One of the best conserved contains a histidine which is one of three residues shown [1] to be involved in the catalytic mechanism of the vertebrate mitochondrial enzyme. This region has been used as a signature pattern.

-Consensus pattern: G-[FYA]-[GA]-H-x-[IV]-x(1,2)-[RKT]-x(2)-D-[PS]-R [H is an active site residue]

[1] Karpusas M., Branchaud B., Remington S.J. Biochemistry 29:2213-2219(1990).

119. clpA_B

Chaperonin clpA/B

CAUTION! This family is a subfamily of the AAA superfamily. The threshold has been set very high to stop overlaps with the AAA superfamily. This entry will be subsumed by AAA in the future.

Number of members: 39

A number of ATP-binding proteins that are are thought to protect cells from extreme stress by controlling the aggregation of denaturation of vital cellular structures have been shown [1,2] to be evolutionary related. These proteins are listed below.

- Escherichia coli clpA, which acts as the regulatory subunit of the ATP-dependent protease clp.
- Rhodopseudomonas blastica clpA homolog.
- Escherichia coli heat shock protein clpB and homologs in other bacteria.
- Bacillus subtilis protein mecB.
- Yeast heat shock protein 104 (gene HSP104), which is vital for tolerance to heat, ethanol and other stresses.
- Neurospora heat shock protein hsp98.
- Yeast mitochondrial heat shock protein 78 (gene HSP78) [3].
- CD4A and CD4b, two highly related tomato proteins that seem to be located in the chloroplast.
- Trypanosoma brucei protein clp.
- Porphyra purpurea chloroplast encoded clpC.

The size of these proteins range from 84 Kd (clpA) to slightly more than 100 Kd (HSP104). They all share two conserved regions of about 200 amino acids that each contains an ATP-binding site. In addition to the ATP-binding A and B motifs there are many parts in these two domains that are also conserved. Two of these regions have been selected as signature patterns. The first signature is located in the first domain, some ten residues to the C-terminal of the ATP-binding B motif. The second pattern is located in the second domain in-

between the ATP-binding A and B motifs.

-Consensus pattern: D-[AI]-[SGA]-N-[LIVMF](2)-K-[PT]-x-L-x(2)-G

-Consensus pattern: R-[LIVMFY]-D-x-S-E-[LIVMFY]-x-E-[KRQ]-x-[STA]-x-[STA]-[KR]-[LIVM]-x-G-[STA]

[1] Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M., Mattick J.S., Dalrymple B., Kuramitsu H., Shiroza T., Foster T., Clark W.P., Ross B., Squires C.L., Maurizi M.R. Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).

[2] Parsell D.A., Sanchez Y., Stitzel J.D., Lindquist S. Nature 353:270-273(1991).

[3] Leonhardt S.A., Fearon K., Danese P.N., Mason T.L. Mol. Cell. Biol. 13:6304-6313(1993).

120. cofilin_ADF

Cofilin/tropomyosin-type actin-binding proteins

[1]
Medline: 97290449
Structure determination of yeast cofilin.
Fedorov AA, Lappalainen P, Fedorov EV, Drubin DG, Almo SC;
Nat Struct Biol 1997;4:366-369.

[2]
Medline: 97290450
Crystal structure of the actin-binding protein actophorin from Acanthamoeba.
Leonard SA, Gittis AG, Petrella EC, Pollard TD, Lattman EE;
Nat Struct Biol 1997;4:369-373.

[3]
Medline: 97420794
F-actin and G-actin binding are uncoupled by mutation of conserved tyrosine residues in maize actin depolymerizing factor.
Jiang CJ, Weeds AG, Khan S, Hussey PJ;
Proc Natl Acad Sci U S A 1997;94:9973-9978.

[4]

Medline: 97357155

Cofilin promotes rapid actin filament turnover in vivo.

Lappalainen P, Drubin DG;

Nature 1997;388:78-82.

Severs actin filaments and binds to actin monomers.

Number of members: 44

Actin-depolymerizing proteins sever actin filaments (F-actin) and/or bind to actin monomers, or G-actin, thus preventing actin-polymerization by sequestering the monomers. The following proteins are evolutionary related and belong to a family of low molecular weight (137 to 166 residues) actin-depolymerizing proteins [1,2,3,4]:

- Cofilin from vertebrates, slime mold and yeast. Cofilin binds to F-actin and acts as a pH-dependent actin-depolymerizing protein.
- Destrin from vertebrates. Destrin binds to G-actin in a pH-independent manner and prevents polymerization.
- Caenorhabditis elegans unc-60.
- Acanthamoeba castellanii actophorin.
- Plants actin depolymerizing factor (ADF).

The most conserved region of these proteins is a twenty amino-acid segment that ends some 30 residues from their C-terminal extremity. This segment has been shown [5] to be important for actin-binding.

-Consensus pattern: P-[DE]-x-[SA]-x-[LIVMT]-[KR]-x-[KR]-M-[LIVM]-[YA]-[STA](3)-x(3)-[LIVMF]-[KR]

[1] Hawkins M., Pope B., MacIver S.K., Weeds A.G. Biochemistry 32:9985-9993(1993).

[2] Iida K., Moriyama K., Matsumoto S., Kawasaki H., Nishida E., Yahara I. Gene 124:115-120(1993).

[3] Quirk S., MacIver S.K., Ampe C., Doberstein S.K., Kaiser D.A., van Damme J., Vandekerckhove J., Pollard T.D. Biochemistry 32:8525-8533(1993).

[4] McKim K.S., Matheson C., Marra M.A., Wakarchuk M.F., Baillie D.L. Mol. Gen. Genet. 242:346-357(1994).

[5] Moriyama K., Yonezawa N., Sakai H., Yahara I., Nishida E. J. Biol. Chem. 267:7240-7244(1992).

121. (Complex 24kd) Respiratory-chain NADH dehydrogenase 24 Kd subunit signature Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complexI or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist inthe chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase).Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 24 Kd (in mammals), which is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a2Fe-2S iron-sulfur cluster. The 24 Kd subunit is nuclear encoded, as aprecursor form with a transit peptide in mammals, and in Neurospora crassa.The 24 Kd subunit is highly similar to [3,4]: - Subunit E of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoE). - Subunit NQO2 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase. A highly conserved region, located in the central section of this subunit containing two conserved cysteines that are probably involved in the binding of the 2Fe-2S center has been selected as a signature pattern.

-Consensus pattern: D-x(2)-F-[ST]-x(5)-C-L-G-x-C-x(2) [GA]-P [The two C's are putative 2Fe-2S ligands]

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

[3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).

[4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

122. copper-bind

Copper binding proteins, plastocyanin/azurin family

Number of members: 70

Blue or 'type-1' copper proteins are small proteins which bind a single copper atom and which are characterized by an intense electronic absorption band near 600 nm [1,2]. The most well known members of this class of proteins are the plant chloroplastic plastocyanins, which exchange electrons with cytochrome c6, and the distantly related bacterial azurins, which exchange electrons with cytochrome c551. This family of proteins also includes all the proteins listed below (references are only provided for recently determined sequences).

- Amicyanin from bacteria such as Methylobacterium extorquens or Thiobacillus versutus that can grow on methylamine. Amicyanin appears to be an electron receptor for methylamine dehydrogenase.
- Auracyanins A and B from Chloroflexus aurantiacus [3]. These proteins can donate electrons to cytochrome c-554.
- Blue copper protein from Alcaligenes faecalis.
- Cupredoxin (CPC) from cucumber peelings [4].
- Cusacyanin (basic blue protein; plantacyanin, CBP) from cucumber.
- Halocyanin from Natrobacterium pharaonis [5], a membrane associated copper-binding protein.
- Pseudoazurin from Pseudomonas.
- Rusticyanin from Thiobacillus ferrooxidans. Rusticyanin is an electron carrier from cytochrome c-552 to the a-type oxidase [6].
- Stellacyanin from the Japanese lacquer tree.
- Umecyanin from horseradish roots.

- Allergen Ra3 from ragweed. This pollen protein is evolutionary related to the above proteins, but seems to have lost the ability to bind copper.

Although there is an appreciable amount of divergence in the sequence of all these proteins, the copper ligand sites are conserved and a pattern which includes two of the ligands (a cysteine and a histidine) has been developed.

-Consensus pattern: [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ] [C and H are copper ligands]

[1] Garret T.P.J., Clingeleffer D.J., Guss J.M., Rogers S.J., Freeman H.C. J. Biol. Chem. 259:2822-2825(1984).

[2] Ryden L.G., Hunt L.T. J. Mol. Evol. 36:41-66(1993).

[3] McManus J.D., Brune D.C., Han J., Sanders-Loehr J., Meyer T.E., Cusanovich M.A., Tollin G., Blankenship R.E. J. Biol. Chem. 267:6531-6540(1992).

[4] Mann K., Schaefer W., Thoenes U., Messerschmidt A., Mehrabian Z., Nalbandyan R. FEBS Lett. 314:220-223(1992).

[5] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

[6] Yano T., Fukumori Y., Yamanaka T. FEBS Lett. 288:159-162(1991).

123. Chaperonins cpn10 signature

Chaperonins [1,2] are proteins involved in the folding of proteins or the assembly of oligomeric protein complexes. They seem to assist other polypeptides in maintaining or assuming conformations which permit their correct assembly into oligomeric structures. They are found in abundance in prokaryotes, chloroplasts and mitochondria. Chaperonins form oligomeric complexes and are composed of two different types of subunits: a 60 Kd protein, known as cpn60 (groEL in bacteria) and a 10 Kd protein, known ascpn10 (groES in bacteria).The cpn10 protein binds to cpn60 in the presence of MgATP and suppresses the ATPase activity of the latter. Cpn10 is a protein of about 100 amino acid residues whose sequence is well conserved in bacteria, vertebrate mitochondriaand plants chloroplast [3,4]. Cpn10 assembles as an heptamer that forms a dome[5]. As a signature pattern for cpn10 , a region located in the N-terminal section of the protein was selected.

Consensus pattern: [LIVMFY]-x-P-[ILT]-x-[DEN]-[KR]-[LIVMFA](3)-[KREQ]-x(8,9)-[SG]-x-[LIVMFY](3)-

Note: this pattern is found twice in the plant chloroplast protein which consist of the tandem repeat of a cpn10 domain

[1] Ellis R.J., van der Vies S.M. Annu. Rev. Biochem. 60:321-347(1991).

[2] Zeilsta-Ryalls J., Fayet O., Georgopoulos C. Annu. Rev. Microbiol. 45:301-325(1991).

[3] Hartman D.J., Hoogenraad N.J., Condron R., Hoj P.B. Proc. Natl. Acad. Sci. U.S.A. 89:3394-3398(1992).

[4] Bertsch U., Soll J., Seetharam R., Viitanen P.V. Proc. Natl. Acad. Sci. U.S.A. 89:8696-8700(1992).

[5] Hunt J.F., Weaver A.J., Landry S.J., Gierasch L., Deisenhofer J. Nature 379:37-45(1996).

124. Chaperonins cpn60 signature (cpn60_TCP1)

Chaperonins [1,2] are proteins involved in the folding of proteins or the assembly of oligomeric protein complexes. Their role seems to be to assist other polypeptides to maintain or assume conformations which permit their correct assembly into oligomeric structures. They are found in abundance in prokaryotes, chloroplasts and mitochondria. Chaperonins form oligomeric complexes and are composed of two different types of subunits: a 60 Kd protein, known as cpn60 (groEL in bacteria) and a 10 Kd protein, known as cpn10 (groES in bacteria).The cpn60 protein shows weak ATPase activity and is a highly conserved protein of about 550 to 580 amino acid residues which has been described by different names in different species: - Escherichia coli groEL protein, which is essential for the growth of the bacteria and the assembly of several bacteriophages. - Cyanobacterial groEL analogues. - Mycobacterium tuberculosis and leprae 65 Kd antigen, Coxiella burnetti heat shock protein B (gene htpB), Rickettsia tsutsugamushi major antigen 58, and Chlamydial 57 Kd hypersensitivity antigen (gene hypB). - Chloroplast RuBisCO subunit binding-protein alpha and beta chains, which bind ribulose bisphosphate carboxylase small and large subunits and are implicated in the assembly of the enzyme oligomer. - Mammalian mitochondrial matrix protein P1 (mitonin or P60). - Yeast HSP60 protein, a mitochondrial assembly factor. As a signature pattern for these proteins, a rather well-conserved region of twelve residues, located in the last third of the cpn60sequence was chosen.

Consensus pattern: A-[AS]-x-[DEQ]-E-x(4)-G-G-[GA]-

[1] Ellis R.J., van der Vies S.M. Annu. Rev. Biochem. 60:321-347(1991).

[2] Zeilsta-Ryalls J., Fayet O., Georgopoulos C. Annu. Rev. Microbiol. 45:301-325(1991).

Chaperonins TCP-1 signatures (cpn60_TCP1)

The TCP-1 protein [1,2] (Tailless Complex Polypeptide 1) was first identified in mice where it is especially abundant in testis but present in all cell types. It has since been found and characterized in many other mammalian species, in Drosophila and in yeast. TCP-1 is a highly conserved protein of about 60 Kd (556 to 560 residues) which participates in a hetero-oligomeric900 Kd double-torus shaped particle [3] with 6 to 8 other different subunits. These subunits, the chaperonin containing TCP-1 (CCT) subunit beta, gamma,delta, epsilon, zeta and eta are evolutionary related to TCP-1 itself [4,5].The CCT is known to act as a molecular chaperone for tubulin, actin and probably some other proteins. The CCT subunits are highly related to archebacterial counterparts: - TF55 and TF56 [6], a molecular chaperone from Sulfolobus shibatae. TF55 has ATPase activity, is known to bind unfolded polypeptides and forms a oligomeric complex of two stacked nine-membered rings. - Thermosome [7], from Thermoplasma acidophilum. The thermosome is composed of two subunits (alpha and beta) and also seems to be a chaperone with ATPase activity. It forms an oligomeric complex of eight-membered rings. The TCP-1 family of proteins are weakly, but significantly [8], related to thecpn60/groEL chaperonin family (see <PDOC00268>).As signature patterns of this family of chaperonins, three conserved regions located in the N-terminal domain were chosen.

Consensus pattern: [RKEL]-[ST]-x-[LMFY]-G-P-x-[GSA]-x-x-K-[LIVMF](2)-

Consensus pattern: [LIVM]-[TS]-[NK]-D-[GA]-[AVNHK]-[TAV]-[LIVM](2)-x(2)-[LIVM]-x-[LIVM]-x-[SNH]-[PQH]-

Consensus pattern: Q-[DEK]-x-x-[LIVMGTA]-[GA]-D-G-T-

[1] Ellis J. Nature 358:191-192(1992).

[2] Nelson R.J., Craig E.A. Curr. Biol. 2:487-489(1992).

[3] Lewis V.A., Hynes G.M., Zheng D., Saibil H., Willison K.R. Nature 358:249-252(1992).

[4] Kubota H., Hynes G., Carne A., Ashworth A., Willison K.R. Curr. Biol. 4:89-99(1994)

[5] Kim S., Willison K.R., Horwich A.L. Trends Biochem. Sci. 20:543-548(1994).

[6] Trent J.D., Nimmesgern E., Wall J.S., Hartl F.U., Horwich A.L. Nature 354:490-493(1991).

[7] Waldmann T., Lupas A., Kellermann J., Peters J., Baumeister W. Biol. Chem. Hoppe-Seyler 376:119-126(1995).

[8] Hemmingsen S.M. Nature 357:650-650(1992).

125. cyclin (Cyclins)

The cyclins include an internal duplication, which is related to that found in TFIIB and the RB protein.

[1]

Medline: 94203808

Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107.

Gibson TJ, Thompson JD, Blocker A, Kouzarides T;

Nucleic Acids Res 1994;22:946-952.

[2]

Medline: 96164440

The crystal structure of cyclin A

Brown NR, Noble MEM, Endicott JA, Garman EF, Wakatsuki S, Mitchell E, Rasmussen B, Hunt T, Johnson LN;

Structure. 1995;3:1235-1247.

Complex of cyclin and cyclin dependant kinase.

[3]

Medline: 96313126

Structural basis of cyclin-dependant kinase activation by phosphorylation.

Russo AA, Jeffrey PD, Pavletich NP;

Nat Struct Biol. 1996;3:696-700.

Cyclins regulate cyclin dependant kinases (CDKs).

The most divergent prosite members have been included. Swiss:P22674 the Uracil-DNA glycosylase 2 is the highest noise and may be related but has not been included.

Number of members: 189

Cyclins [1,2,3] are eukaryotic proteins which play an active role in controlling nuclear cell division cycles. Cyclins, together with the p34 (cdc2) or cdk2 kinases, form the Maturation Promoting Factor (MPF). There are two main groups of cyclins:

- G2/M cyclins, essential for the control of the cell cycle at the G2/M (mitosis) transition. G2/M cyclins accumulate steadily during G2 and are abruptly destroyed as cells exit from mitosis (at the end of the M-phase).
- G1/S cyclins, essential for the control of the cell cycle at the G1/S (start) transition.

In most species, there are multiple forms of G1 and G2 cyclins. For example, in vertebrates, there are two G2 cyclins, A and B, and at least three G1 cyclins, C, D, and E.

A cyclin homolog has also been found in herpesvirus saimiri [4].

The best conserved region is in the central part of the cyclins' sequences, known as the 'cyclin-box'. From this, a 32 residue pattern has been derived.

-Consensus pattern: R-x(2)-[LIVMSA]-x(2)-[FYWS]-[LIVM]-x(8)-[LIVMFC]-x(4)-[LIVMFYA]-x(2)-[STAGC]-[LIVMFYQ]-x-[LIVMFYC]-[LIVMFY]-D-[RKH]-[LIVMFYW]

[1] Nurse P. Nature 344:503-508(1990).
[2] Norbury C., Nurse P. Curr. Biol. 1:23-24(1991).
[3] Lew D.J., Reed S.I. Trends Cell Biol. 2:77-81(1992).
[4] Nicholas J., Cameron K.R., Honess R.W. Nature 355:362-365(1992).

126. Cystatin domain

This is a very diverse family. Attempts to define separate subfamilies have failed. Typically, either the N-terminal or C-terminal end is very divergent. But splitting into two domains would make very short families. Cathelicidins are related to this family but have not been included. Number of members: 147

Inhibitors of cysteine proteases [1,2,3], which are found in the tissues and body fluids of animals, in the larva of the worm Onchocerca volvulus [4], as well as in plants, can be grouped into three distinct but related families:

- Type 1 cystatins (or stefins), molecules of about 100 amino acid residues with neither disulfide bonds nor carbohydrate groups.
- Type 2 cystatins, molecules of about 115 amino acid residues which contain one or two disulfide loops near their C-terminus.
- Kininogens, which are multifunctional plasma glycoproteins.

They are the precursor of the active peptide bradykinin and play a role in blood coagulation by helping to position optimally prekallikrein and factor XI next to factor XII. They are also inhibitors of cysteine proteases. Structurally, kininogens are made of three contiguous type-2 cystatin domains, followed by an additional domain (of variable length) which contains the sequence of bradykinin. The first of the three cystatin domains seems to have lost its inhibitory activity.

In all these inhibitors, there is a conserved region of five residues which has been proposed to be important for the binding to the cysteine proteases. The consensus pattern starts one residue before this conserved region.

-Consensus pattern: [GSTEQKRV]-Q-[LIVT]-[VAF]-[SAGQ]-G-x-[LIVMNK]-x(2)-[LIVMFY]-x-[LIVMFYA]-[DENQKRHSIV]

[1] Barrett A.J. Trends Biochem. Sci. 12:193-196(1987).
[2] Rawlings N.D., Barrett A.J. J. Mol. Evol. 30:60-71(1990).
[3] Turk V., Bode W. FEBS Lett. 285:213-219(1991).
[4] Lustigman S., Brotman B., Huima T., Prince A.M. Mol. Biochem. Parasitol. 45:65-76(1991).

127. cytochrome_c (Cytochrome c)
The Pfam entry does not include all prosite members.
The cytochrome 556 and cytochrome c' families are not included.
Number of members: 259

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of

the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

-Consensus pattern: C-{CPWHF}-{CPWR}-C-H-{CFYW}

[1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

128. (DAGKa) Diacylglycerol kinase accessory domain (presumed)

Diacylglycerol (DAG) is a second messenger that acts as a protein kinase C activator. This domain is assumed to be an accessory domain: its function is unknown.

[1] Sakane F, Yamada K, Kanoh H, Yokoyama C, Tanabe T, Nature 1990;344:345-348.[2] Sakane F, Imai S, Kai M, Wada I, Kanoh H, J Biol Chem 1996;271:8394-8401.[3] Schaap D, de Widt J, van der Wal J, Vandekerckhove J, van, Damme J, Gussow D, Ploegh HL, van Blitterswijk WJ, van der, Bend RL, FEBS Lett 1990;275:151-158. [4] Kanoh H, Yamada K, Sakane F, Trends Biochem Sci 1990;15:47-50.

129. (DAGKc) Diacylglycerol kinase catalytic domain (presumed)

Diacylglycerol (DAG) is a second messenger that acts as a protein kinase C activator. The catalytic domain is assumed from the finding of bacterial homologues.

[1] Sakane F, Yamada K, Kanoh H, Yokoyama C, Tanabe T, Nature 1990;344:345-348. [2] Sakane F, Imai S, Kai M, Wada I, Kanoh H, J Biol Chem 1996;271:8394-8401. [3] Schaap D, de Widt J, van der Wal J, Vandekerckhove J, van, Damme J, Gussow D, Ploegh HL, van Blitterswijk WJ, van der, Bend RL, FEBS Lett 1990;275:151-158. [4] Kanoh H, Yamada K, Sakane F, Trends Biochem Sci 1990;15:47-50.

130. D-amino acid oxidases signature(DAO)

D-amino acid oxidase (EC 1.4.3.3) (DAMOX or DAO) is an FAD flavoenzyme that catalyzes the oxidation of neutral and basic D-amino acids into their corresponding keto acids. DAOs have been characterized and sequenced in fungi and vertebrates where they are known to be

located in the peroxisomes. D-aspartate oxidase (EC 1.4.3.1) (DASOX) [1] is an enzyme, structurally related to DAO, which catalyzes the same reaction but is active only toward dicarboxylic D-amino acids. In DAO, a conserved histidine has been shown [2] to be important for the enzyme's catalytic activity. The conserved region around this residue has been developed as a signature pattern for these enzymes.

Consensus pattern: [LIVM](2)-H-[NHA]-Y-G-x-[GSA](2)-x-G-x(5)-G-x-A [H is a probable active site residue]o-

[1] Negri A., Ceciliani F., Tedeschi G., Simonic T., Ronchi S. J. Biol. Chem. 267:11865-11871(1992).

[2] Miyano M., Fukui K., Watanabe F., Takahashi S., Tada M., Kanashiro M., Miyake Y. J. Biochem. 109:171-177(1991).

131. DEAD and DEAH box families ATP-dependent helicases signatures

A number of eukaryotic and prokaryotic proteins have been characterized [1,2,3] on the basis of their structural similarity. They all seem to be involved in ATP-dependent, nucleic-acid unwinding. Proteins currently known to belong to this family are: - Initiation factor eIF-4A. Found in eukaryotes, this protein is a subunit of a high molecular weight complex involved in 5'cap recognition and the binding of mRNA to ribosomes. It is an ATP-dependent RNA-helicase. - PRP5 and PRP28. These yeast proteins are involved in various ATP-requiring steps of the pre-mRNA splicing process. - Pl10, a mouse protein expressed specifically during spermatogenesis. - An3, a Xenopus putative RNA helicase, closely related to Pl10. - SPP81/DED1 and DBP1, two yeast proteins probably involved in pre-mRNA splicing and related to Pl10. - Caenorhabditis elegans helicase glh-1. - MSS116, a yeast protein required for mitochondrial splicing. - SPB4, a yeast protein involved in the maturation of 25S ribosomal RNA. - p68, a human nuclear antigen. p68 has ATPase and DNA-helicase activities in vitro. It is involved in cell growth and division. - Rm62 (p62), a Drosophila putative RNA helicase related to p68. - DBP2, a yeast protein related to p68. - DHH1, a yeast protein. - DRS1, a yeast protein involved in ribosome assembly. - MAK5, a yeast protein involved in maintenance of dsRNA killer plasmid. - ROK1, a yeast protein. - ste13, a fission yeast protein. - Vasa, a Drosophila protein important for oocyte formation and specification of embryonic posterior structures. - Me31B, a Drosophila maternally expressed protein of

unknown function. - dbpA, an Escherichia coli putative RNA helicase. - deaD, an Escherichia coli putative RNA helicase which can suppress a mutation in the rpsB gene for ribosomal protein S2. - rhlB, an Escherichia coli putative RNA helicase. - rhlE, an Escherichia coli putative RNA helicase. - srmB, an Escherichia coli protein that shows RNA-dependent ATPase activity. It probably interacts with 23S ribosomal RNA. - Caenorhabditis elegans hypothetical proteins T26G10.1, ZK512.2 and ZK686.2. - Yeast hypothetical protein YHR065c. - Yeast hypothetical protein YHR169w. - Fission yeast hypothetical protein SpAC31A2.07c. - Bacillus subtilis hypothetical protein yxiN. All these proteins share a number of conserved sequence motifs. Some of them are specific to this family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases `superfamily' [4,E1]. One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins [3,5,6,E1]. Proteins currently known to belong to this subfamily are: - PRP2, PRP16, PRP22 and PRP43. These yeast proteins are all involved in various ATP-requiring steps of the pre-mRNA splicing process. - Fission yeast prh1, which my be involved in pre-mRNA splicing. - Male-less (mle), a Drosophila protein required in males, for dosage compensation of X chromosome linked genes. - RAD3 from yeast. RAD3 is a DNA helicase involved in excision repair of DNA damaged by UV light, bulky adducts or cross-linking agents. Fission yeast rad15 (rhp3) and mammalian DNA excision repair protein XPD (ERCC-2) are the homologs of RAD3. - Yeast CHL1 (or CTF1), which is important for chromosome transmission and normal cell cycle progression in G(2)/M. - Yeast TPS1. - Yeast hypothetical protein YKL078w. - Caenorhabditis elegans hypothetical proteins C06E1.10 and K03H1.2. - Poxviruses' early transcription factor 70 Kd subunit which acts with RNA polymerase to initiate transcription from early gene promoters. - I8, a putative vaccinia virus helicase. - hrpA, an Escherichia coli putative RNA helicase. Signature patterns for both subfamilies were developed.

Consensus pattern: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN

Consensus pattern: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR]

Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see the relevant entry <PDOC00017

[1] Schmid S.R., Linder P. Mol. Microbiol. 6:283-292(1992).

[2] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[3] Wassarman D.A., Steitz J.A. Nature 349:463-464(1991).

[4] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[5] Harosh I., Deschavanne P. Nucleic Acids Res. 19:6331-6331(1991).

[6] Koonin E.V., Senkevich T.G. J. Gen. Virol. 73:989-993(1992).

132. (DHBP_synthase) 3,4-dihydroxy-2-butanone 4-phosphate synthase

3,4-Dihydroxy-2-butanone 4-phosphate is biosynthesized from ribulose 5-phosphate and serves as the biosynthetic precursor for the xylene ring of riboflavin. Sometimes found as a bifunctional enzyme with GTP_cyclohydro2.

Richter G, Krieger C, Volk R, Kis K, Ritz H, Gotze E, Bacher A, Methods Enzymol 1997;280:374-382.

133. (DHDPS) Dihydrodipicolinate synthetase signatures

Dihydrodipicolinate synthetase (EC 4.2.1.52) (DHDPS) [1] catalyzes, in higher plants chloroplast and in many bacteria (gene dapA), the first reaction specific to the biosynthesis of lysine and of diaminopimelate. DHDPS is responsible for the condensation of aspartate semialdehyde and pyruvate by aping-pong mechanism in which pyruvate first binds to the enzyme by forming a Schiff-base with a lysine residue. Three other proteins are structurally related to DHDPS and probably also act via a similar catalytic mechanism: - Escherichia coli N-acetylneuraminate lyase (EC 4.1.3.3) (gene nanA), which catalyzes the condensation of N-acetyl-D-mannosamine and pyruvate to form N-acetylneuraminate. - Rhizobium meliloti protein mosA [3], which is involved in the biosynthesis of the rhizopine 3-o-methyl-scyllo-inosamine. - Escherichia coli hypothetical protein yjhH. Two signature patterns for these enzymes were developed . The first one is centered on highly conserved region in the N-terminal part of these proteins. The second signature contains a lysine residue which has been shown, in Escherichia coli dapA [2], to be the one that forms a Schiff-base with the substrate.

Consensus pattern: [GSA]-[LIVM]-[LIVMFY]-x(2)-G-[ST]-[TG]-G-E-[GASNF]-x(6)- [EQ]

-

Consensus pattern: Y-[DNS]-[LIVMFA]-P-x(2)-[ST]-x(3)-[LIVMG]-x(13,14)-[LIVM]- x-[SGA]-[LIVMF]-K-[DEQAF]-[STAC] [K is involved in Schiff-base formation]-

[1] Kaneko T., Hashimoto T., Kumpaisal R., Yamada Y. J. Biol. Chem. 265:17451-17455(1990).

[2] Laber B., Gomis-Rueth F.-X., Romao M.J., Huber R. Biochem. J. 288:691-695(1992).

[3] Murphy P.J., Trenz S.P., Grzemski W., de Bruijn F.J., Schell J. J. Bacteriol. 175:5193-5204 (1993).

134. (DHOdehase) Dihydroorotate dehydrogenase signatures

Dihydroorotate dehydrogenase (EC 1.3.3.1) (DHOdehase) catalyzes the fourth step in the de novo biosynthesis of pyrimidine, the conversion of dihydroorotate into orotate. DHOdehase is a ubiquitous FAD flavoprotein. In bacteria (gene pyrD), DHOdease is located on the inner side of the cytosolic membrane. In some yeasts, such as in Saccharomyces cerevisiae (gene URA1), it is a cytosolic protein while in other eukaryotes it is found in the mitochondria [1]. The sequence of DHOdease is rather well conserved and two signature patterns were developed specific to this enzyme. The first corresponds to a region in the N-terminal section of the enzyme while the second is located in the C-terminal section and seems to be part of the FAD-binding domain.

Consensus pattern[GS]-x(4)-[GK]-[GSTA]-[LIVFSTA]-[GT]-x(3)-[NQR]-x-G-[NHY]-x(2)-P-[RT]

Consensus pattern[LIVM](2)-[GSA]-x-G-G-[IV]-x-[STGDN]-x(3)-[ACV]-x(6)-G-A

[1] Nagy M., Lacroute F., Thomas D. Proc. Natl. Acad. Sci. U.S.A. 89:8966-8970(1992).

135. (DMRL_synthase) 6,7-dimethyl-8-ribityllumazine synthase

136. (DNA_methylase) C-5 cytosine-specific DNA methylases signatures

C-5 cytosine-specific DNA methylases (EC 2.1.1.73) (C5 Mtase) are enzymes that specifically methylate the C-5 carbon of cytosines in DNA [1,2,3]. Such enzymes are found

in the proteins described below. - As a component of type II restriction-modification systems in prokaryotes and some bacteriophages. Such enzymes recognize a specific DNA sequence where they methylate a cytosine. In doing so, they protect DNA from cleavage by type II restriction enzymes that recognize the same sequence. The sequences of a large number of type II C-5 Mtases are known. - In vertebrates, there are a number of C-5 Mtases that methylate CpG dinucleotides. The sequence of the mammalian enzyme is known.C-5 Mtases share a number of short conserved regions. Two of them were selected. The first is centered around a conserved Pro-Cys dipeptide in which the cysteine has been shown [4] to be involved in the catalytic mechanism; it appears to form a covalent intermediate with the C6 position of cytosine. The second region is located at the C-terminal extremity in type-II enzymes

Consensus pattern: [DENKS]-x-[FLIV]-x(2)-[GSTC]-x-P-C-x(2)-[FYWLIM]-S [C is the active site residue]-

Consensus pattern: [RKQGTF]-x(2)-G-N-[STAG]-[LIVMF]-x(3)-[LIVMT]-x(3)-[LIVM]-x(3)-[LIVM]-

[1] Posfai J., Bhagwat A.S., Roberts R.J. Gene 74:261-263(1988).

[2] Kumar S., Cheng X., Klimasauskas S., Mi S., Posfai J., Roberts R.J., Wilson G.G. Nucleic Acids Res. 22:1-10(1994).

[3] Lauster R., Trautner T.A., Noyer-Weidner M. J. Mol. Biol. 206:305-312(1989).

[4] Chen L., McMillan A.M., Chang W., Ezak-Nipkay K., Lane W.S., Verdine G.L. Biochemistry 30:11018-11025(1991).

137. (DNAphotolyase) DNA photolyases class 2 signatures

Deoxyribodipyrimidine photolyase (EC <u>4.1.99.3</u>) (DNA photolyase) [1,2] is a DNArepair enzyme. It binds to UV-damaged DNA containing pyrimidine dimers and, upon absorbing a near-UV photon (300 to 500 nm), breaks the cyclobutane ring joining the two pyrimidines of the dimer. DNA photolyase is an enzyme that requires two choromophore-cofactors for its activity: a reduced FADH2 and either 5,10-methenyltetrahydrofolate (5,10-MTFH) or an oxidized 8-hydroxy-5-deazaflavin (8-HDF) derivative (F420). The folate or deazaflavin chromophore appears to function as an antenna, while the FADH2 chromophore is thought to be responsible for electron transfer. On the basis of sequence similarities[3] DNA

photolyases can be grouped into two classes. The second class contains enzymes from Myxococcus xanthus, methanogenic archaebacteria, insects, fish and marsupial mammals. It is not yet known what second cofactor is bound to class 2 enzymes. There are a number of conserved sequence regions in all known class 2 DNAphotolyases, especially in the C-terminal part. Two of these regions were selected as signature patterns.

Consensus pattern: F-x-E-E-x-[LIVM](2)-R-R-E-L-x(2)-N-F-

Consensus pattern: G-x-H-D-x(2)-W-x-E-R-x-[LIVM]-F-G-K-[LIVM]-R-[FY]-M-N-

[1] Sancar G.B., Sancar A. Trends Biochem. Sci. 12:259-261(1987).

[2] Jorns M.S. Biofactors 2:207-211(1990).

[3] Yasui A., Eker A.P.M., Yasuhira S., Yajima H., Kobayashi T., Takao M., Oikawa A. EMBO J. 13:6143-6151(1994).

(DNAphotolyase2) DNA photolyases class 1 signatures

Deoxyribodipyrimidine photolyase (EC 4.1.99.3) (DNA photolyase) [1,2] is a DNA repair enzyme. It binds to UV-damaged DNA containing pyrimidine dimers and ,upon absorbing a near-UV photon (300 to 500 nm), breaks the cyclobutane ring joining the two pyrimidines of the dimer. DNA photolyase is an enzyme that requires two choromophore-cofactors for its activity: a reduced FADH2 and either 5,10-methenyltetrahydrofolate (5,10-MTFH) or an oxidized 8-hydroxy-5-deazaflavin (8-HDF) derivative (F420). The folate or deazaflavin chromophore appears to function as an antenna, while the FADH2 chromophore is thought to be responsible for electron transfer. On the basis of sequence similarities[3] DNA photolyases can be grouped into two classes. The first class contains enzymes from Gram-negative and Gram-positive bacteria, the halophilic archaebacteria Halobacterium halobium, fungi and plants. Class 1 enzymes bind either 5,10-MTHF (E.coli, fungi, etc.) or 8-HDF (S.griseus, H.halobium).This family also includes Arabidopsis cryptochromes 1 (CRY1) and 2 (CRY2),which are blue light photoreceptors that mediate blue light-induced gene expression. There are a number of conserved sequence regions in all known class 1 DNA photolyases, especially in the C-terminal part. Two of these regions were selected as signature patterns

Consensus pattern: T-G-x-P-[LIVM](2)-D-A-x-M-[RA]-x-[LIVM]-

Consensus pattern: [DN]-R-x-R-[LIVM](2)-x-[STA](2)-F-[LIVMFA]-x-K-x-L-x(2,3)- W-[KRQ]-

[1] Sancar G.B., Sancar A. Trends Biochem. Sci. 12:259-261(1987).

[2] Jorns M.S. Biofactors 2:207-211(1990).

[3] Yasui A., Eker A.P.M., Yasuhira S., Yajima H., Kobayashi T., Takao M., Oikawa A. EMBO J. 13:6143-6151(1994).

[4] Lin C., Ahmad M., Cashmore A.R. Plant J. 10:893-902(1996).

138. (DNA_pol_A)

DNA polymerase family A signature

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarities a number of DNA polymerases have been grouped together [1,2,3] under the designation of DNA polymerase family A. The polymerases that belong to this family are listed below.

- Escherichia coli and various other bacterial polymerase I (gene polA).
- Thermus aquaticus Taq polymerase.
- Bacteriophage sp01 polymerase.
- Bacteriophage sp02 polymerase.
- Bacteriophage T5 polymerase.
- Bacteriophage T7 polymerase.
- Mycobacteriophage L5 polymerase.
- Yeast mitochondrial polymerase gamma (gene MIP1).

Five regions of similarity are found in all the above polymerases. One of these conserved regions, known as 'motif B' [1], is located in a domain which, in Escherichia coli polA, has been shown to bind deoxynucleotide triphosphate substrates; it contains a conserved tyrosine which has been shown, by photo- affinity labelling, to be in the active site; a conserved lysine, also part of this motif, can be chemically labelled, using pyridoxal phosphate. This conserved region was used as a signature for this family of DNA polymerases.

Consensus patternR-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA] Sequences known to belong to this class detected by the pattern ALL.

[1] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Eng. 3:461-467(1990).
[2] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
[3] Braithwaite D.K., Ito J. Nucleic Acids Res. 21:787-802(1993).

139. DNA_pol_viral_C
DNA polymerase (viral) C-terminal domain
Number of members: 128

140. (DNA_topoisoII)
DNA topoisomerase II signature
DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

```
<----------------About-1400-residues---------------------->

[----------Protein 39-*------][----Protein 52----]        Phage T4
[----------gyrB-------*-----][--------gyrA----------------]  Prokaryote II
                                                  Archaebacteria
[----------parE-------*-----][--------parD----------------]  Prokaryote IV
```

[--------------------*---------------------------------] Eukaryote and ASF

'*': Position of the pattern.

As a signature pattern for this family of proteins, a region that contains a highly conserved pentapeptide was selected. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

Consensus pattern[LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the pattern ALL.

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
[2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
[3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

141. (DSPc) Tyrosine specific protein phosphatases signature and profiles

Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below: Soluble PTPases. - PTPN1 (PTP-1B). - PTPN2 (T-cell PTPase; TC-PTP). - PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1- like domain (see <PDOC00566>) and could act at junctions between the membrane and cytoskeleton. - PTPN5 (STEP). - PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup. - PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP). - PTPN8 (70Z-PEP). - PTPN9 (MEG2). - PTPN12 (PTP-G1; PTP-P19). - Yeast PTP1. - Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway. - Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis. - Fission yeast pyp3 which contributes to the dephosphorylation of cdc2. - Yeast CDC14 which may

be involved in chromosome segregation. - Yersinia virulence plasmid PTPAses (gene yopH). - Autographa californica nuclear polyhedrosis virus 19 Kd PTPase.Dual specificity PTPases. - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185. - DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues. - DUSP3 (VHR). - DUSP4 (HVH2). - DUSP5 (HVH3). - DUSP6 (Pyst1; MKP-3). - DUSP7 (Pyst2; MKP-X). - Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3. - Yeast YVH1. - Vaccinia virus H1 PTPase; a dual specificity phosphatase. Receptor PTPases. Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminalcatalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectintype III (FN-III) repeats, immunoglobulin-like domains, MAM domains orcarbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not. In the following table, the domain structure of known receptor PTPases is shown: Extracellular Intracellular ------------------ ------------- Ig FN-3 CAH MAM PTPaseLeukocyte common antigen (LCA) (CD45) 0 2 0 0 2Leukocyte antigen related (LAR) 3 8 0 0 2 Drosophila DLAR 3 9 0 0 2Drosophila DPTP 2 2 0 0 2PTP-alpha (LRP) 0 0 0 0 2PTP-beta 0 16 0 0 1PTP-gamma 0 1 1 0 2PTP-delta 0 >7 0 0 2 PTP-epsilon 0 0 0 0 2PTP-kappa 1 4 0 1 2PTP-mu 1 4 0 1 2PTP-zeta 0 1 1 0 2PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important. A signature pattern for PTPase domains was derived centered on the active site cysteine. There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily

Consensus pattern: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue]-

[1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991).
[2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992).
[3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991).

[4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989).

[5] Hunter T. Cell 58:1013-1016(1989).

142. (DUF10) Uncharacterized protein family UPF0076 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Goat antigen UK114, a human homolog and the rat corresponding protein which is known as perchloric acid soluble protein (PSP1). PSP1 [2] may inhibit an initiation stage of cell-free protein synthesis. - Mouse heat-responsive protein HRSP12. - Yeast chromosome V hypothetical protein YER057c. - Yeast chromosome IX hypothetical protein YIL051c. - Caenorhabditis elegans hypothetical protein C23G10.2. - Escherichia coli hypothetical protein ycdK. - Escherichia coli hypothetical protein yhaR. - Escherichia coli hypothetical protein yjgF and HI0719, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yoaB. - Bacillus subtilis hypothetical protein yabJ. - Haemophilus influenzae hypothetical protein HI1627. - Helicobacter pylori hypothetical protein HP0944. - Lactococcus lactis aldR. - Myxococcus xanthus dfrA. - Synechocystis strain PCC 6803 hypothetical protein slr0709. - Rhizobium strain NGR234 symbiotic plasmid hypothetical protein y4sK. - Pyrococcus horikoshii hypothetical protein PH0854.These are small proteins of around 15 Kd whose sequence is highly conserved.As a signature pattern, a well conserved region located in the C-terminal part of these proteins was selected.

Consensus pattern: [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)-[GSAKR]-x-[LMVA]-x(5,8)-[LIVM]-E-[MI]-

[1] Bairoch A. Unpublished observations (1995).

[2] Oka T., Tsuji H., Noda C., Sakai K., Hong Y.-M., Suzuki I., Munoz S., Natori Y. J. Biol. Chem. 270:30060-30067(1995).

143. (DUF3)Domain of Unknown Function 3

Domain apparently occurring exclusively in eubacteria. Unknown function.

144. (DUF6) Integral membrane protein

This family includes many hypothetical membrane proteins of unknown function. Many of the proteins contain two copies of the aligned region.

145. (DUF7) Integral membrane protein

This family includes many hypothetical membrane proteins of unknown function. Swiss:P14502 has been implicated in resistance to ethidium bromide.

146. (DapB) Dihydrodipicolinate reductase signature

Dihydrodipicolinate reductase (EC 1.3.1.26) catalyzes the second step in the biosynthesis of diaminopimelic acid and lysine, the NAD or NADP-dependent reduction of 2,3-dihydrodipicolinate into 2,3,4,5-tetrahydrodipicolinate. This enzyme is present in bacteria (gene dapB) and higher plants. As a signature pattern the best conserved region in this enzyme was selected. It is located in the central section and is part of the substrate-binding region [1].

Consensus pattern: E-[IV]-x-E-x-H-x(3)-K-x-D-x-P-S-G-T-A-

[1] Scapin G., Blanchard J.S., Sacchettini J.C. Biochemistry 34:3502-3512(1995).

147. DedA family

This family combines the DedA related proteins and YIAN/YGIK family. Members of this family are not functionally characterised. These proteins contain multiple predicted transmembrane regions.

148. DegT/DnrJ/EryC1/StrS family

The members of this family exhibit some characteristics of the sensor protein of two-component signal transduction systems, however none of the members show any sequence similarity to these protein kinases. The members of this family do have the typical helix-turn-helix motif of DNA binding proteins.

[1] Stutzman-Engwall KJ, Otten SL, Hutchinson CR, J Bacteriol 1992;174:144-154.

149. (Desaturase) Fatty acid desaturases signatures

Fatty acid desaturases (EC 1.14.99.-) are enzymes that catalyze the insertion of a double bond at the delta position of fatty acids. There seems to be two distinct families of fatty acid desaturases which do not seem to be evolutionary related. Family 1 is composed of: - Stearoyl-CoA desaturase (SCD) (EC 1.14.99.5) [1]. SCD is a key regulatory enzyme of unsaturated fatty acid biosynthesis. SCD introduces a cis double bond at the delta(9) position of fatty acyl-CoA's such as palmitoleoyl- and oleoyl-CoA. SCD is a membrane-bound enzyme that is thought to function as a part of a multienzyme complex in the endoplasmic reticulum of vertebrates and fungi. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected, this region is rich in histidine residues and in aromatic residues. Family 2 is composed of: - Plants stearoyl-acyl-carrier-protein desaturase (EC 1.14.99.6) [2], these enzymes catalyze the introduction of a double bond at the delta(9) position of steraoyl-ACP to produce oleoyl-ACP. This enzyme is responsible for the conversion of saturated fatty acids to unsaturated fatty acids in the synthesis of vegetable oils. - Cyanobacteria desA [3] an enzyme that can introduce a second cis double bond at the delta(12) position of fatty acid bound to membranes glycerolipids. DesA is involved in chilling tolerance; the phase transition temperature of lipids of cellular membranes being dependent on the degree of unsaturation of fatty acids of the membrane lipids. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected.

Consensus pattern: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y-

Consensus pattern: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]- [DE]-

[1] Kaestner K.H., Ntambi J.M., Kelly T.J. Jr., Lane M.D. J. Biol. Chem. 264:14755-14761(1989).

[2] Shanklin J., Somerville C.R. Proc. Natl. Acad. Sci. U.S.A. 88:2510-2514(1991).

[3] Wada H., Gombos Z., Murata N. Nature 347:200-203(1990).

150. Dihydroorotase signatures

Dihydroorotase (EC 3.5.2.3) (DHOase) catalyzes the third step in the de novo biosynthesis of pyrimidine, the conversion of ureidosuccinic acid (N-carbamoyl-L-aspartate) into dihydroorotate. Dihydroorotase binds a zinc ion which is required for its catalytic activity [1]. In bacteria, DHOase is a dimer of identical chains of about 400 amino-acid residues (gene pyrC). In higher eukaryotes, DHOase is part of a large multi-functional protein known as 'rudimentary' in Drosophila and CAD in mammals and which catalyzes the first three steps of pyrimidine biosynthesis [2]. The DHOase domain is located in the central part of this polyprotein. In yeasts, DHOase is encoded by a monofunctional protein (gene URA4). However, a defective DHOase domain [3] is found in a multifunctional protein (gene URA2)that catalyzes the first two steps of pyrimidine biosynthesis. The comparison of DHOase sequences from various sources shows [4] that there are two highly conserved regions. The first located in the N-terminal extremity contains two histidine residues suggested [3] to be involved in binding the zinc ion. The second is found in the C-terminal part. Signature patterns for both regions have been developed. Allantoinase (EC 3.5.2.5) is the enzyme that hydrolyzes allantoin intoallantoate. In yeast (gene DAL1) [5], it is the first enzyme in the allanto indegradation pathway; in amphibians [6] and fish it catalyzes the second step in the degradation of uric acid. The sequence of allantoinase is evolutionary related to that of DHOases.

Consensus pattern: D-[LIVMFYWSAP]-H-[LIVA]-H-[LIVF]-[RN]-x-[PGANF] [The two H's are probable zinc ligands]-

Consensus pattern: [GA]-[ST]-D-x-A-P-H-x(4)-K-

[1] Brown D.C., Collins K.D. J. Biol. Chem. 266:1597-1604(1991).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[3] Souciet J.-L., Nagy M., Le Gouar M., Lacroute F., Potier S. Gene 79:59-70(1989).

[4] Guyonvarch A., Nguyen-Juilleret M., Hubert J.-C., Lacroute F. Mol. Gen. Genet. 212:134-141(1988).

[5] Buckholz R.G., Cooper T.G. Yeast 7:913-923(1991).

[6] Hayashi S., Jain S., Chu R., Alvares K., Xu B., Erfurth F., Usuda N., Rao M.S., Reddy S.K., Noguchi T., Reddy J.K., Yeldandi A.Y. J. Biol. Chem. 269:12269-12276(1994).

151. dnaJ domains signatures and profile

The prokaryotic heat shock protein dnaJ interacts with the chaperone hsp70-like dnaK protein [1]. Structurally, the dnaJ protein consists of an N- terminal conserved domain (called 'J' domain) of about 70 amino acids, a glycine-rich region ('G' domain') of about 30 residues, a central domain containing four repeats of a CXXCXGXG motif ('CRR' domain) and a C-terminal region of 120 to 170 residues. Such a structure is shown in the following schematic representation:

+------------+-+-------+-----+-----------+------------------------------+ | N-terminal | | Gly-R | | CXXCXGXG | C-terminal | +------------+-+-------+-----+-----------+--------------------- -----------+

It has been shown [2] that the 'J' domain as well as the 'CRR' domain are also found in other prokaryotic and eukaryotic proteins which are listed below.

a) Proteins containing both a 'J' and a 'CRR' domain:

- Yeast protein MAS5/YDJ1 which seems to be involved in mitochondrial protein import.
- Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding.
- Yeast protein SCJ1, involved in protein sorting.
- Yeast protein XDJ1.
- Plants dnaJ homologs (from leek and cucumber).
- Human HDJ2, a dnaJ homolog of unknown function.
- Yeast hypothetical protein YNL077w.

b) Proteins containing a 'J' domain without a 'CRR' domain:

- Rhizobium fredii nolC, a protein involved in cultivar-specific nodulation of soybean.
- Escherichia coli cbpA [3], a protein that binds curved DNA.
- Yeast protein SEC63/NPL1, important for protein assembly into the endoplasmic reticulum and the nucleus.
- Yeast protein SIS1, required for nuclear migration during mitosis.
- Yeast protein CAJ1.
- Yeast hypothetical protein YFR041c.
- Yeast hypothetical protein YIR004w.
- Yeast hypothetical protein YJL162c.

- Plasmodium falciparum ring-infected erythrocyte surface antigen (RESA). RESA, whose function is not known, is associated with the membrane skeleton of newly invaded erythrocytes.
- Human HDJ1.
- Human HSJ1, a neuronal protein.
- Drosophila cysteine-string protein (csp).

A signature pattern for the 'J' domain was developed, based on conserved positions in the C-terminal half of this domain. A pattern for the 'CRR' domain, based on the first two copies of that motif was also developed. A profile for the 'J' domain was also developed.

Consensus pattern: [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI]-

Consensus pattern: C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G-

[1] Cyr D.M., Langer T., Douglas M.G. Trends Biochem. Sci. 19:176-181(1994).
[2] Bork P., Sander C., Valencia A., Bukau B. Trends Biochem. Sci. 17:129-129(1992).
[3] Ueguchi C., Kaneda M., Yamada H., Mizuno T. Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).

152.

153. Dwarfin

This family known as the dwarfins also includes the drosophila protein MAD. The N-terminus of MAD can bind to DNA [2].

[1] Yingling JM, Das P, Savage C, Zhang M, Padgett RW, Wang XF, Proc Natl Acad Sci U S A 1996;93:8940-8944. [2] Kim J, Johnson K, Chen HJ, Carroll S, Laughon A, Nature 1997;388:304-308.

154. Dynein light chain type 1 signature

Dynein is a multisubunit microtubule-dependent motor enzyme that acts as the force generating protein of eukaryotic cilia and flagella. The cytoplasmic isoform of dynein acts as

a motor for the intracellular retrograde motility of vesicles and organelles along microtubules. Dynein is composed of a number of ATP-binding large subunits, intermediate size subunits and small subunits. Among the small subunits, there is a family [1,2] of highly conserved proteins which consist of: - Chlamydomonas reinhardtii flagellar outer arm dynein 8 Kd and 11 Kd light chains. - Higher eukaryotes cytoplasmic dynein light chain 1. - Yeast cytoplasmic dynein light chain 1 (gene DYN2 or SLC1). - Caenorhabditis elegans hypothetical dynein light chains M18.2 and T26A5.9.These proteins are have from 89 to 120 amino acids. As a signature pattern, A highly conserved region was selected.

Consensus pattern: H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E -

[1] King S.M., Patel-King R.S. J. Biol. Chem. 270:11445-11452(1995).
[2] Dick T., Ray K., Salz H.K., Chia W. Mol. Cell. Biol. 16:1966-1977(1996).

155. dUTPase

dUTPase hydrolyzes dUTP to dUMP and pyrophosphate.

[1] Cedergren-Zeppezauer ES, Larsson G, Nyman PO, Dauter Z, Wilson KS, Nature 1992;355:740-743. [2] Mol CD, Harris JM, McIntosh EM, Tainer JA, Structure 1996;4:1077-1092.

156. (dCMP cyt deam) Cytidine and deoxycytidylate deaminases zinc-binding region signature

Cytidine deaminase (EC 3.5.4.5) (cytidine aminohydrolase) catalyzes the hydrolysis of cytidine into uridine and ammonia while deoxycytidylatedeaminase (EC 3.5.4.12) (dCMP deaminase) hydrolyzes dCMP into dUMP. Both enzymes are known to bind zinc and to require it for their catalytic activity[1,2]. These two enzymes do not share any sequence similarity with the exception of a region that contains three conserved histidine and cysteine residues which are thought to be involved in the binding of the catalytic zincion. Such a region is also found in other proteins [3,4]: - Yeast cytosine deaminase (EC 3.5.4.1) (gene FCY1) which transforms cytosine into uracil. - Mammalian apolipoprotein B mRNA editing protein, responsible for the postranscriptional editing of a CAA codon into a UAA (stop) codon in the APOB mRNA. - Riboflavin biosynthesis protein ribG, which converts 2,5-

diamino-6- (ribosylamino)-4(3H)-pyrimidinone 5'-phosphate into 5-amino-6- (ribosylamino)-2,4(1H,3H)-pyrimidinedione 5'-phosphate. - Bacillus cereus blasticidin-S deaminase (EC 3.5.4.23), which catalyzes the deamination of the cytosine moiety of the antibiotics blasticidin S, cytomycin and acetylblasticidin S. - Bacillus subtilis protein comEB. This protein is required for the binding and uptake of transforming DNA. - Bacillus subtilis hypothetical protein yaaJ. - Escherichia coli hypothetical protein yfhC. - Yeast hypothetical protein YJL035c. A signature pattern for this zinc-binding region was derived.

Consensus pattern: [CH]-[AGV]-E-x(2)-[LIVMFGAT]-[LIVM]-x(17,33)-P-C-x(2,8)-C-x(3)-[LIVM] [The C's and H are zinc ligands

[1] Yang C., Carlow D., Wolfenden R., Short S.A. Biochemistry 31:4168-4174(1992).
[2] Moore J.T., Silversmith R.E., Maley G.F., Maley F. J. Biol. Chem. 268:2288-2291(1993).
[3] Reizer J., Buskirk S., Bairoch A., Reizer A., Saier M.H. Jr. Protein Sci. 3:853-856(1994).
[4] Bhattacharya S., Navaratnam N., Morrison J.R., Scott J., Taylow W.R. Trends Biochem. Sci. 19:105-106(1994).

157. Dehydrins signatures

A number of proteins are produced by plants that experience water-stress. Water-stress takes place when the water available to a plant falls below a critical level. The plant hormone abscisic acid (ABA) appears to modulate the response of plant to water-stress. Proteins that are expressed during water-stress are called dehydrins [1,2] or LEA group 2 proteins [3]. The proteins that belong to this family are listed below.

- Arabidopsis thaliana XERO 1, XERO 2 (LTI30), RAB18, ERD10 (LTI45) ERD14 and COR47.
- Barley dehydrins B8, B9, B17, and B18.
- Cotton LEA protein D-11.
- Craterostigma plantagineum dessication-related proteins A and B.
- Maize dehydrin M3 (RAB-17).
- Pea dehydrins DHN1, DHN2, and DHN3.
- Radish LEA protein.
- Rice proteins RAB 16B, 16C, 16D, RAB21, and RAB25.

- Tomato TAS14.
- Wheat dehydrin RAB 15 and cold-shock protein cor410, cs66 and cs120.

Dehydrins share a number of structural features. One of the most notable features is the presence, in their central region, of a continuous run of five to nine serines followed by a cluster of charged residues. Such a region has been found in all known dehydrins so far with the exception of pea dehydrins. A second conserved feature is the presence of two copies of alysine-rich octapeptide; the first copy is located just after the cluster of charged residues that follows the poly-serine region and the second copy is found at the C-terminal extremity. Signature patterns for both regions were derived.

Consensus pattern: S(5)-[DE]-x-[DE]-G-x(1,2)-G-x(0,1)-[KR](4

Consensus pattern: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G-

[1] Close T.J., Kortt A.A., Chandler P.M. Plant Mol. Biol. 13:95-108(1989).

[2] Robertson M., Chandler P.M. Plant Mol. Biol. 19:1031-1044(1992).

[3] Dure L. III, Crouch M., Harada J., Ho T.-H. D., Mundy J., Quatrano R., Thomas T., Sung Z.R. Plant Mol. Biol. 12:475-486(1989).

158. (deoR) Bacterial regulatory proteins, deoR family signature

The many bacterial transcription regulation proteins which bind DNA through a helix-turn-helix' motif can be classified into subfamilies on the basis of sequence similarities. One of these subfamilies groups the following proteins[1,2]: - accR, the Agrobacterium tumefaciens plasmid pTiC58 repressor of opine catabolism and conjugal transfer. - agaR, the Escherichia coli aga operon putative repressor. - deoR, the Escherichia coli deoxyribose operon repressor. - fucR, the Escherichia coli L-fucose operon activator. - gatR, the Escherichia coli galactitol operon repressor. - glpR, the Escherichia coli glycerol-3-phosphate regulon repressor. - gutR (or srlR), the Escherichia coli glucitol operon repressor. - iolR, from Bacillus subtilis. - lacR, the streptococci lactose phosphotransferase system repressor. - spoIIID, the Bacillus subtilis transcription regulator of the sigK gene. - yfjR, an Escherichia coli hypothetical protein. - ygbI, an Escherichia coli hypothetical protein. - yihW, an Escherichia coli hypothetical protein. - yjfQ, an Escherichia coli hypothetical protein. - yjhJ, an Escherichia coli hypothetical protein. The 'helix-turn-helix' DNA-binding motif of these proteins is located in the N-terminal part of the sequence. The pattern used to detect these proteins starts fourteen residues before the HTH motif and ends one residue after it.

Consensus pattern: R-x(3)-[LIVM]-x(3)-[LIVM]-x(16,17)-[STA]-x(2)-T-[LIVMA]- [RH]-[KRNA]-D-[LIVMF]-

[1] von Bodman S., Hayman G.T., Farrand S.K. Proc. Natl. Acad. Sci. U.S.A. 89:643-647(1992).

[2] Bairoch A. Unpublished observations (1993).

159. dsrm

Double-stranded RNA binding motif

[1] Burd CG, Dreyfuss G; Medline: 94310455, Conserved structures and diversity of functions of RNA-binding proteins. Science 1994;265:615-621.

Sequences gathered for seed by HMM_iterative_training Putative motif shared by proteins that bind to dsRNA. At least some DSRM proteins seem to bind to specific RNA targets. Exemplified by Staufen, which is involved in localization of at least five different mRNAs in the early Drosophila embryo. Also by interferon-induced protein kinase in humans, which is part of the cellular response to dsRNA.

Number of members: 116

160. Dynamin family signature

Dynamin [1,2] is a microtubule-associated force-producing protein of 100 Kd which is involved in the production of microtubule bundles and which is able to bind and hydrolyze GTP. Dynamin is structurally related to the following proteins: - Drosophila shibire protein (gene shi) [3]. Shibire is, very probably, the Drosophila cognate of mammalian dynamin. It seems to provide the motor for vesicular transport during endocytosis. - Yeast vacuolar sorting protein VPS1 (or SPO15) [4], a protein which could also be involved in microtubule-associated motility. - Yeast protein MGM1 [5], which is required for mitochondrial genome maintenance. - Yeast protein DNM1, which is involved in endocytosis. - Interferon induced Mx proteins [6,7]. Interferon alpha or beta induce the synthesis of a family of closely related

proteins. Most of these proteins are known to confer resistance to influenza viruses and/or rhabdoviruses on transfected mammalian cell in culture. The three motifs found in all GTP-binding proteins are located in the N-terminal part of these proteins. The signature pattern that was developed for these proteins is based on a highly conserved region downstream of the ATP/GTP-binding motif 'A' (P-loop) (see <PDOC00017>).-

Consensus pattern: L-P-[RK]-G-[STN]-[GN]-[LIVM]-V-T-R-

[1] Vallee R.B., Shpetner H.S. Annu. Rev. Biochem. 59:909-932(1990).
[2] Obar R.A., Collins C.A., Hammarback J.A., Shpetner H.S., Vallee R.B. Nature 347:256-261(1990).
[3] van der Bliek A., Meyerowitz E.M. Nature 351:411-414(1991).
[4] Rothman J.H., Raymond C.K., Gilbert T., O'Hara P.J., Stevens T.H. Cell 61:1063-1074(1990).
[5] Jones B.A., Fangman W.L. Genes Dev. 6:380-389(1992).
[6] Arnheiter H., Meier E. New Biol. 2:851-857(1990).
[7] Staeheli P., Pitossi F., Pavlovic J. Trends Cell Biol. 3:268-272(1993).

161. (dynamin_2) Dynamin central region

This region lies between the GTPase domain, see dynamin, and the pleckstrin homology (PH) domain.

162. E1-E2 ATPases phosphorylation site

E1-E2 ATPases (also known as P-type) are cation transport ATPases which form an aspartyl phosphate intermediate in the course of ATP hydrolysis. ATPases which belong to this family are listed below [1,2,3]. - Fungal and plant plasma membrane (H+) ATPases [reviewed in 4]. - Vertebrate (Na+, K+) ATPases (sodium pump) [reviewed in 5,6]. - Gastric (K+, H+) ATPases (proton pump). - Calcium (Ca++) ATPases (calcium pump) from the sarcoplasmic reticulum (SR), the endoplasmic reticulum (ER) and the plasma membrane. - Copper (Cu++) ATPases (copper pump) which are involved in two human genetic disorders: Menkes syndrome and Wilson disease [7]. - Bacterial potassium (K+) ATPases. - Bacterial cadmium efflux (Cd++) ATPases [reviewed in 8]. - Bacterial magnesium (Mg++) ATPases. - A

probable cation ATPase from Leishmania. - fixI, a probable cation ATPase from Rhizobium meliloti, involved in nitrogen fixation. The region around the phosphorylated aspartate residue is perfectly conserved in all these ATPases and can be used as a signature pattern.

Consensus pattern: D-K-T-G-T-[LI]-[TI] [D is phosphorylated]

[1] Green N.M., McLennan D.H. Biochem. Soc. Trans. 17:819-822(1989).
[2] Green N.M. Biochem. Soc. Trans. 17:970-972(1989).
[3] Fagan M.J., Saier M.H. Jr. J. Mol. Evol. 38:57-99(1994).
[4] Serrano R. Biochim. Biophys. Acta 947:1-28(1988).
[5] Fambrough D.M. Trends Neurosci. 11:325-328(1988).
[6] Sweadner K.J. Biochim. Biophys. Acta 988:185-220(1989).
[7] Bull P.C., Cox D.W. Trends Genet. 10:246-251(1994).
[8] Silver S., Nucifora G., Chu L., Misra T.K. Trends Biochem. Sci. 14:76-80(1989).

163. E1_N
E1 Protein, N terminal domain
Number of members: 90

164. (E1_dehydrog) Dehydrogenase E1 component

This family uses thiamine pyrophosphate as a cofactor. This family includes pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and 2-oxoisovalerate dehydrogenase.

165. (ECH) Enoyl-CoA hydratase/isomerase signature

Enoyl-CoA hydratase (EC 4.2.1.17) (ECH) [1] and 3-2trans-enoyl-CoA isomerase(EC 5.3.3.8) (ECI) [2] are two enzymes involved in fatty acid metabolism. ECH catalyzes the hydratation of 2-trans-enoyl-CoA into 3-hydroxyacyl-CoA and ECI shifts the 3- double bond of the intermediates of unsaturated fatty acid oxidation to the 2-trans position. Most eukaryotic cells have two fatty-acid beta-oxidation systems, one located in mitochondria and the other in peroxisomes. In mitochondria, ECH and ECI are separate yet structurally related monofunctional enzymes. Peroxisomes contain a trifunctional enzyme [3] consisting of an N-

terminal domain that bears both ECH and ECI activity, and a C-terminal domain responsible for 3-hydroxyacyl-CoA dehydrogenase (HCDH) activity. In Escherichia coli (gene fadB) and Pseudomonas fragi (gene faoA), ECH and ECI are also part of a multifunctional enzyme which contains both a HCDH and a3-hydroxybutyryl-CoA epimerase domain [4].A number of other proteins have been found to be evolutionary related to the ECH/ECI enzymes or domains: - 3-hydroxbutyryl-coa dehydratase (EC 4.2.1.55) (crotonase), a bacterial enzyme involved in the butyrate/butanol-producing pathway. - Naphthoate synthase (EC 4.1.3.36) (DHNA synthetase) (gene menB) [5], a bacterial enzyme involved in the biosynthesis of menaquinone (vitamin K2). DHNA synthetase converts O-succinyl-benzoyl-CoA (OSB-CoA) to 1,4-dihydroxy- 2-naphthoic acid (DHNA). - 4-chlorobenzoate dehalogenase (EC 3.8.1.6) [6], a Pseudomonas enzyme which catalyzes the conversion of 4-chlorobenzoate-CoA to 4-hydroxybenzoate-CoA. - A Rhodobacter capsulatus protein of unknown function (ORF257) [7]. - Bacillus subtilis putative polyketide biosynthesis proteins pksH and pksI. - Escherichia coli carnitine racemase (gene caiD) [8]. - Escherichia coli hypothetical protein ygfG. - Yeast hypothetical protein YDR036c.As a signature pattern for these enzymes, a conserved region richin glycine and hydrophobic residues was selected.

Consensus pattern: [LIVM]-[STA]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-[DQHP]-[LIVMFY]-

[1] Minami-Ishii N., Taketani S., Osumi T., Hashimoto T. Eur. J. Biochem. 185:73-78(1989).

[2] Mueller-Newen G., Stoffel W. Biol. Chem. Hoppe-Seyler 372:613-624(1991).

[3] Palosaari P.M., Hiltunen J.K. J. Biol. Chem. 265:2446-2449(1990).

[4] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:4937-4937(1990).

[5] Driscoll J.R., Taber H.W. J. Bacteriol. 174:5063-5071(1992).

[6] Babbitt P.C., Kenyon G.L., Matin B.M., Charest H., Sylvestre M., Scholten J.D., Chang K.-H., Liang P.-H., Dunaway-Mariano D. Biochemistry 31:5594-5604(1992).

[7] Beckman D.L., Kranz R.G. Gene 107:171-172(1991).

[8] Eichler K., Bourgis F., Buchet A., Kleber H.-P., Mandrand-Berthelot M.-A. Mol. Microbiol. 13:775-786(1994).

166. (EF1BD) Elongation factor 1 beta/beta'/delta chain signatures

Eukaryotic elongation factor 1 (EF-1) is responsible for the GTP-dependent binding of aminoacyl-tRNAs to the ribosomes [1]. EF-1 is composed of four subunits: the alpha chain which binds GTP and aminoacyl-tRNAs, the gamma chain that probably plays a role in anchoring the complex to other cellular components and the beta and delta (or beta') chains. The beta and delta chains are highly similar proteins that both stimulate the exchange of GDP bound to the alpha chain for GTP [2]. The beta and delta chains are hydrophilic proteins of around 23 to 31 Kd. Their C-terminal part seems important for the nucleotide exchange activity, while the N-terminal section is probably involved in the interaction with the gamma chain. Two signature patterns for this family of proteins were developed. The first corresponds to an acidic region in the central section; the second, to the C-terminal extremity of these proteins

Consensus pattern: [DE]-[DEG]-[DE](2)-[LIVMF]-D-L-F-G-
Consensus pattern: [IV]-Q-S-x-D-[LIVM]-x-A-[FWM]-[NQ]-K-[LIVM]-

[1] Riis B., Rattan I.S., Clark B.F.C., Merrick W.C. Trends Biochem. Sci. 15:420-424(1990).
[2] van Damme H.T.F., Amons R., Karssies R., Timmers C.J., Janssen G.M.C., Moeller W. Biochim. Biophys. Acta 1050:241-247(1990).

167. (EF1G_domain) Elongation factor 1 gamma, conserved domain

168. (EFG_C) Elongation factor G C-terminus

This family is always found associated with GTP_EFTU. This family includes the carboxyl terminal regions of Elongation factor G, elongation factor 2 and some tetracycline resistance proteins.

169. (EFP) Elongation factor P signature

Elongation factor P (EF-P) [1] is a prokaryotic protein translation factor required for efficient peptide bond synthesis on 70S ribosomes from fMet-tRNAfMet. EF-P is a protein of 21 Kd. It is evolutionary related to yeiP, an hypothetical protein from Escherichia coli. As a

signature pattern, a conserved region located in the C-terminal part of these proteins was selected.

Consensus pattern: K-x-[AV]-x(4)-G-x(2)-[LIV]-x-V-P-x(2)-[LIV]-x(2)-G-

[1] Aoki H., Adams S.-L., Turner M.A., Ganoza M.C. Biochimie 79:7-11(1997).

170. (EF TS) Elongation factor Ts signatures

In prokaryotes elongation factor Ts (EF-Ts) is a component of the elongation cycle of protein biosynthesis. It associates with the EF-Tu.GDP complex and induces the exchange of GDP to GTP, it remains bound to the aminoacyl-tRNA.EF-Tu.GTP complex up to the GTP hydrolysis stage on the ribosome [1].EF-Ts is also a component of the chloroplast protein biosynthetic machinery and is encoded in the genome of some algal chloroplast [2]. It is also present in mitochondria [3]. As signature patterns for EF-Ts, two conserved regions located in the N-terminal part of the protein have been selected.

Consensus pattern: L-R-x(2)-T-[GSDNQ]-x-[GS]-[LIVMF]-x(0,1)-[DENKAC]-x-K-[KRNEQS]-A-L-

Consensus pattern: E-[LIVM]-[NV]-[SCV]-[QE]-T-D-F-V-[SA]-[KRN]-

[1] Bubunenko M.G., Kireeva M.L., Gudkov A.T. Biochimie 74:419-425(1992).
[2] Kostrzewa M., Zetsche K. Plant Mol. Biol. 23:67-76(1993).
[3] Xin H., Woriax V.L., Burkhart W.A., Spremulli L.L. J. Biol. Chem. 270:17243-17249(1995).

171. (EMP24_GP25L) emp24/gp25L/p24 family

Members of this family are implicated in bringing cargo forward from the ER and binding to coat proteins by their cytoplasmic domains. Number of members: 30

Paccaud JP, Thomas DY, Bergeron JJ, Nilsson T, J Cell Biol 1998;140:751-765.

172. ENV_polyprotein

ENV polyprotein (coat polyprotein)

Number of members: 224

173. (ERG4_ERG24) Ergosterol biosynthesis ERG4/ERG24 family signatures

Two fungal enzymes involved in ergosterol biosynthesis and which act by reducing double bonds in precursors of ergosterol have been shown to be evolutionary related [1]. These are C-14 sterol reductase (gene ERG24 in budding yeast and erg3 in Neurospora Crassa) and C-24(28) sterol reductase (gene ERG4 in budding yeast and sts1 in fission yeast). Their sequences are also highly related to that of chicken lamin B receptor, which is thought to anchor the lamina to the inner nuclear membrane. These proteins are highly hydrophobic and seem to contain seven or eight transmembrane regions. As signature patterns, two conserved regions were selected. The first one is apparently located in a loop between the fourth and fifth transmembrane regions and the second is in the C-terminal section.

Consensus pattern: G-x(2)-[LIVM]-[YH]-D-x-[FYW]-x-G-x(2)-L-N-P-R -

Consensus pattern: [LIVM](2)-H-R-x(2)-R-D-x(3)-C-x(2)-K-Y-G-

[1] Lai M.H., Bard M., Pierson C.A., Alexander J.F., Goebl M., Carter G.T., Kirsch D.R. Gene 140:41-49(1994).

174. (ERM) Ezrin/radixin/moesin family

This family of proteins contain a band 4.1 domain (Band_41), at their amino terminus. This family represents the rest of these proteins.

[1] Yonemura S, Hirao M, Doi Y, Takahashi N, Kondo T, Tsukita S, J Cell Biol 1998;140:885-895.

175. ER lumen protein retaining receptor signatures

Proteins that reside in the lumen of the endoplasmic reticulum (ER) contain aC-terminal tetrapeptide (generally K-D-E-L or H-D-E-L) that serves as a signal for their retrieval (retrograde transport) from subsequent compartments of the secretory pathway. The signal is recognized by a receptor molecule that is believed to cycle between the cis side of the Golgi

apparatus and the ER [1].This protein is known as the ER lumen protein retaining receptor or also as the 'KDEL receptor'. It has been characterized in a variety of species, including fungi (gene ERD2), plants, Plasmodium, Drosophila and mammals. In mammals two highly related forms of the receptor are known. Structurally, the receptor is a protein of about 220 residues that seems to contain seven transmembrane regions [2]. The N-terminal part (3 residues) is oriented toward the lumen while the C-terminal tail (about 12 residues) is cytoplasmic. There are three lumenal and three cytoplasmic loops. Two signature patterns for these receptors were developed. The first pattern corresponds to the C-terminal half of the first cytoplasmic loop as well as most of the second transmembrane domain. The second pattern is a perfectly conserved decapeptide that corresponds to the central part of the fifth transmembrane domain.

Consensus pattern: G-I-S-x-[KR]-x-Q-x-L-[FY]-x-[LIV](2)-F-x(2)-R-Y-
Consensus pattern: L-E-[SA]-V-A-I-[LM]-P-Q-L-

[1] Pelham H.R.B. Curr. Opin. Cell Biol. 3:585-591(1991).
[2] Townsley F.M., Wilson D.W., Pelham H.R.B. EMBO J. 12:2821-2829(1993).

176. (ETF_beta) Electron transfer flavoprotein beta-subunit signature
The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria. The beta subunit of ETF is a protein of about 28 Kd which is structurally related to the bacterial nitrogen fixation protein fixA which could play a role in a redox process and feed electrons to ferredoxin. Other related proteins are: - Escherichia coli hypothetical protein ydiQ. - Escherichia coli hypothetical protein ygcR.As a signature pattern for these proteins, a conserved region which is located in the central section was selected.

Consensus pattern: [IVA]-x-[KR]-x(2)-[DE]-[GD]-[GDE]-x(1,2)-[EQ]-x-[LIV]- x(4)-P-x-[LIVM](2)-[TAC]-

[1] Finocchiaro G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990).

[2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).

177. Endonuclease III signatures

Escherichia coli endonuclease III (EC 4.2.99.18) (gene nth) [1] is a DNA repair enzyme that acts both as a DNA N-glycosylase, removing oxidized pyrimidines from DNA, and as an apurinic/apyrimidinic (AP) endonuclease, introducing a single-strand nick at the site from which the damaged base was removed. Endonuclease III is an iron-sulfur protein that binds a single 4Fe-4Scluster. The 4Fe-4S cluster does not seem to be important for catalytic activity, but is probably involved in the proper positioning of the enzyme along the DNA strand [2].Endonuclease III is evolutionary related to the following proteins: - Fission yeast endonuclease III homolog (gene nth1) [3]. - Escherichia coli and related protein DNA repair protein mutY, which is an adenine glycosylase. MutY is a larger protein (350 amino acids) than endonuclease III (211 amino acids). - Micrococcus luteus ultraviolet N-glycosylase/AP lyase which initiates repair at cis-syn pyrimidine dimers. - ORF10 in plasmid pFV1 of the thermophilic archaebacteria Methanobacterium thermoformicicum [4]. Restriction methylase m.MthTI, which is encoded by this plasmid, generates 5-methylcytosine which is subject to deamination resulting in G-T mismatches. This protein could correct these mismatches. - Yeast hypothetical protein YAL015c. - Fission yeast hypothetical protein SpAC26A3.02. - Caenorhabditis elegans hypothetical protein R10E4.5. - Methanococcus jannaschii hypothetical protein MJ0613.The 4Fe-4S cluster is bound by four cysteines which are all located in a 17amino acid region at the C-terminal end of endonuclease III. A similar region is also present in the central section of mutY and in the C-terminus of ORF10and of the Micrococcus UV endonuclease. The 4Fe-4S cluster region does not exist in YAL015c. Two signature patterns for these proteins were developed: the first corresponds to the core of the iron-sulfur binding domain, the second corresponds to the best conserved region in the catalytic core of these enzymes.

Consensus pattern: C-x(3)-[KRS]-P-[KRAGL]-C-x(2)-C-x(5)-C [The four C's are 4Fe-4S ligands]-

Consensus pattern: [GST]-x-[LIVMF]-P-x(5)-[LIVMW]-x(2,3)-[LI]-[PAS]-G-V-[GA]- x(3)-[GAC]-x(3)-[LIVM]-x(2)-[SALV]-[LIVMFYW]-[GANK]-

[1] Kuo C.-F., McRee D., Fisher C.L., O'Handley S.F., Cunnigham R.P., Tainer J.A. Science 258:434-440(1992).

[2] Thomson A.J. Curr. Biol. 3:173-174(1993).

[3] Roldan-Arjona T., Anselmino C., Lindahl T. Nucleic Acids. Res. 3307-3312(1996).

[4] Noelling J., van Eeden F.J.M., Eggen R.I.L., de Vos W.M. Nucleic Acids Res. 20:6501-6507(1992).

178. (Epimerase) NAD dependent epimerase/dehydratase family

This family of proteins utilize NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions.

[1] Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM, Biochemistry 1997;36:6294-6304.

179. Exonuclease

This family includes a variety of exonuclease proteins, such as ribonuclease T and the epsilon subunit of DNA polymerase III.

[1] Koonin EV, Deutscher MP, Nucleic Acids Res 1993;21:2521-2522.

180. ENTH

ENTH domain

[1] Kay BK, Yamabhai M, Wendland B, Emr SD; Medline: 99156083, Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery. Protein Sci 1999;8:435-438.

The ENTH (Epsin N-terminal homology) domain is found in proteins involved in endocytosis and cytoskeletal machinery. The function of the ENTH domain is unknown.

Number of members: 29

181. (eIF-1A) Eukaryotic initiation factor 1A signature

Eukaryotic translation initiation factor 1A (eIF-1A) [1] (formerly known aseiF-4C) is a protein that seems to be required for maximal rate of protein biosynthesis. It enhances ribosome dissociation into subunits and stabilizesthe binding of the initiator Met-tRNA to 40S ribosomal subunits.eIF-1A is a hydrophilic protein of about 15 to 17 Kd. Archaebacteria also seem to possess a eIF-1A homolog. As a signature pattern, a conserved region in the central section of these proteins was selected.

Consensus pattern: [IM]-x-G-x-[GS]-[KRH]-x(4)-[CL]-x-D-G-x(2)-R-x(2)-[RH]-I- x-G

[1] Wei C.-L., Kainuma M., Hershey J.W.B. J. Biol. Chem. 270:22788-22794(1995).

182. (eIF-5A) Eukaryotic initiation factor 5A hypusine signature

Eukaryotic initiation factor 5A (eIF-5A) (formerly known as eIF-4D) [1,2] is a small protein whose precise role in the initiation of protein synthesis is not known. It appears to promote the formation of the first peptide bond. eIF-5Aseems to be the only eukaryotic protein to contain an hypusine residue. Hypusine is derived from lysine by the post-translational addition of a butylamino group (from spermidine) to the epsilon-amino group of lysine. The hypusine group is essential to the function of eIF-5A.A hypusine-containing protein has been found in archaebacteria such as Sulfolobus acidocaldarius or Methanococcus jannaschii; this protein is highlysimilar to eIF-5A and could play a similar role in protein biosynthesis. The signature developed for eIF-5A is centered around the hypusine residue.

Consensus pattern: [PT]-G-K-H-G-x-A-K [The first K is modified to hypusine]

[1] Park M.H., Wolff E.C., Folk J.E. Biofactors 4:95-104(1993).

[2] Schnier J., Schwelberger H.G., Smit-McBride Z., Kang H.A., Hershey J.W.B. Mol. Cell. Biol. 11:3105-3114(1991).

183. (efhand) S-100/ICaBP type calcium binding protein signature

S-100 are small dimeric acidic calcium and zinc-binding proteins [1] abundant in the brain. They have two different types of calcium-binding sites: a low affinity one with a special

structure and a 'normal' EF-hand type high affinity site. The vitamin-D dependent intestinal calcium-binding proteins (ICaBP or calbindin 9 Kd) also belong to this family of proteins, but it does not form dimers. In the past years the sequences of many new members of this family have been determined (for reviews see [2,3,4]); in most cases the function of these proteins is not yet known, although it is becoming clearthat they are involved in cell growth and differentiation, cell cycle regulation and metabolic control. These proteins are: - Calcyclin (Prolactin receptor associated protein (PRA); clatropin; 2a9; 5B10; S100A6). - Calpactin I light chain (p10; p11; 42c; S100A10). - Calgranulin A (cystic fibrosis antigen (CFAg); MIF related protein 8 (MRP- 8); p8; S100A8). - Calgranulin B (MIF related protein 14 (MRP-14); p14; S100A9). - Calgranulin C. - Calgizzarin (S100C). - Placental calcium-binding protein (CAPL) (18a2; peL98; 42a; p9K; MTS1; metastatin; S100A4). - Protein S-100D (S100A5). - Protein S-100E (S100A3). - Protein S-100L (CAN19; S100A2). - Placental protein S-100P (S100E). - Psoriasin (S100A7). - Chemotactic cytokine CP-10 [5]. - Protein MRP-126 [6]. - Trichohyalin [7]. This is a large intermediate filament-associated protein that associates with keratin intermediate filaments (KIF); it contains a S- 100 type domain in its N-terminal extremity. A number of these proteins are known to bind calcium while others are not (p10for example). Our EF-hand detecting pattern will fail to pick those proteins which have lost their calcium-binding properties. A pattern was developed which unambiguously picks up proteins belonging to this family. This pattern spans the region of the EF-hand high affinity site but makes no assumptions on the calcium-binding properties of this site.

Consensus pattern: [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x- [ES]-[FYVC]-x(2)-[LIVMFS]-[LIVMF]

[1] Baudier J. (In) Calcium and Calcium Binding proteins, Gerday C., Bollis L., Giller R., Eds., pp102-113, Springer Verlag, Berlin, (1988).

[2] Moncrief N.D., Kretsinger R.H., Goodman M. J. Mol. Evol. 30:522-562(1990).

[3] Kligman D., Hilt D.C. Trends Biochem. Sci. 13:437-443(1988).

[4] Schaefer B.W., Wicki R., Engelkamp D., Mattei M.-G., Heizmann C.W. Genomics 25:638-643(1995).

[5] Lackmann M., Cornish C.J., Simpson R.J., Moritz R.L., Geczy C.L. J. Biol. Chem. 267:7499-7504(1992).

[6] Nakano T., Graf T. Oncogene 7:527-534(1992).

[7] Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D., Steinert P.M., J. Biol. Chem. 268:12164-12176(1993).

EF-hand calcium-binding domain

Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand [1 to 5]. This type of domain consists of a twelve residue loop flanked on both side by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Listed below are the proteins which are known to contain EF-hand regions. For each type of protein the total number of EF-hand regions known or supposed to exist is indicated between parenthesis. This number does not include regions which clearly have lost their calcium-binding properties, or the atypical low-affinity site (which spans thirteen residues) found in the S-100/ ICaBP family of proteins [6].

- Aequorin and Renilla luciferin binding protein (LBP) (Ca=3).
- Alpha actinin (Ca=2). - Calbindin (Ca=4).
- Calcineurin B subunit (protein phosphatase 2B regulatory subunit) (Ca=4).
- Calcium-binding protein from Streptomyces erythraeus (Ca=3?).
- Calcium-binding protein from Schistosoma mansoni (Ca=2?).
- Calcium-binding proteins TCBP-23 and TCBP-25 from Tetrahymena thermophila (Ca=4?). - Calcium-dependent protein kinases (CDPK) from plants (Ca=4).
- Calcium vector protein from amphoxius (Ca=2).
- Calcyphosin (thyroid protein p24) (Ca=4?).
- Calmodulin (Ca=4, except in yeast where Ca=3).
- Calpain small and large chains (Ca=2). - Calretinin (Ca=6).
- Calcyclin (prolactin receptor associated protein) (Ca=2).
- Caltractin (centrin) (Ca=2 or 4).
- Cell Division Control protein 31 (gene CDC31) from yeast (Ca=2?).
- Diacylglycerol kinase (EC 2.7.1.107) (DGK) (Ca=2).
- FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) from

mammals (Ca=1). - Fimbrin (plastin) (Ca=2).

- Flagellar calcium-binding protein (1f8) from Trypanosoma cruzi (Ca=1 or 2).
- Guanylate cyclase activating protein (GCAP) (Ca=3).
- Inositol phospholipid-specific phospholipase C isozymes gamma-1 and delta-1 (Ca=2) [10]. - Intestinal calcium-binding protein (ICaBPs) (Ca=2).
- MIF related proteins 8 (MRP-8 or CFAG) and 14 (MRP-14) (Ca=2).
- Myosin regulatory light chains (Ca=1). - Oncomodulin (Ca=2).
- Osteonectin (basement membrane protein BM-40) (SPARC) and proteins that contains an 'osteonectin' domain (QR1, matrix glycoprotein SC1) (see the entry <PDOC00535>) (Ca=1). - Parvalbumins alpha and beta (Ca=2).
- Placental calcium-binding protein (18a2) (nerve growth factor induced protein 42a) (p9k) (Ca=2).
- Recoverins (visinin, hippocalcin, neurocalcin, S-modulin) (Ca=2 to 3).
- Reticulocalbin (Ca=4). - S-100 protein, alpha and beta chains (Ca=2).
- Sarcoplasmic calcium-binding protein (SCPs) (Ca=2 to 3).
- Sea urchin proteins Spec 1 (Ca=4), Spec 2 (Ca=4?), Lps-1 (Ca=8).
- Serine/threonine protein phosphatase rdgc (EC 3.1.3.16) from Drosophila (Ca=2) - Sorcin V19 from hamster (Ca=2). - Spectrin alpha chain (Ca=2).
- Squidulin (optic lobe calcium-binding protein) from squid (Ca=4).
- Troponins C; from skeletal muscle (Ca=4), from cardiac muscle (Ca=3), from arthropods and molluscs (Ca=2).

There has been a number of attempts [7,8] to develop patterns that pick-up EF-hand regions, but these studies were made a few years ago when not so many different families of calcium-binding proteins were known. Therefore a new pattern was developed which takes into account all published sequences. This pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic.

-Consensus pattern: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW]
-Note: positions 1 (X), 3 (Y) and 12 (-Z) are the most conserved.
-Note: the 6th residue in an EF-hand loop is, in most cases a Gly, but the number of exceptions to this 'rule' has gradually increased and therefore the pattern should include all

the different residues which have been shown to exist in this position in functional Ca-binding sites.

-Note: the pattern will, in some cases, miss one of the EF-hand regions in some proteins with multiple EF-hand domains.

[1] Kawasaki H., Kretsinger R.H. Protein Prof. 2:305-490(1995).[2] Kretsinger R.H. Cold Spring Harbor Symp. Quant. Biol. 52:499-510(1987).

[3] Moncrief N.D., Kretsinger R.H., Goodman M. J. Mol. Evol. 30:522-562(1990).

[4] Nakayama S., Moncrief N.D., Kretsinger R.H. J. Mol. Evol. 34:416-448(1992).

[5] Heizmann C.W., Hunziker W. Trends Biochem. Sci. 16:98-103(1991).

[6] Kligman D., Hilt D.C. Trends Biochem. Sci. 13:437-443(1988).

[7] Strynadka N.C.J., James M.N.G.
Annu. Rev. Biochem. 58:951-98(1989).

[8] Haiech J., Sallantin J. Biochimie 67:555-560(1985).

[9] Chauvaux S., Beguin P., Aubert J.-P., Bhat K.M., Gow L.A., Wood T.M., Bairoch A. Biochem. J. 265:261-265(1990).

[10] Bairoch A., Cox J.A. FEBS Lett. 269:454-456(1990).

184. Enolase signature

Enolase (EC <u>4.2.1.11</u>) is a glycolytic enzyme that catalyzes the dehydration of2-phospho-D-glycerate to phosphoenolpyruvate [1]. It is a dimeric enzyme that requires magnesium both for catalysis and stabilizing the dimer. Enolase is probably found in all organisms that metabolize sugars. In vertebrates, there are three different tissue-specific isozymes: alpha present in most tissues, beta in muscles and gamma found only in nervous tissues. Tau-crystallin, one of the major lens proteins in some fish, reptiles and birds, has been shown [2] to be evolutionary related to enolase. As a signature pattern for enolase, the best conserved region was selected, it is located in the C-terminal third of the sequence.-

Consensus pattern: [LIV](3)-K-x-N-Q-I-G-[ST]-[LIV]-[ST]-[DE]-[STA]

[1] Lebioda L., Stec B., Brewer J.M. J. Biol. Chem. 264:3685-3693(1989).

[2] Wistow G., Piattigorsky J. Science 236:1554-1556(1987).

185. (F-actin_cap_A) F-actin capping protein alpha subunit signatures

The F-actin capping protein binds in a calcium-independent manner to the fast growing ends of actin filaments (barbed end) thereby blocking the exchange of subunits at these ends. Unlike gelsolin and severin this protein does not sever actin filaments. The F-actin capping protein is a heterodimer composed of two unrelated subunits: alpha and beta.The alpha subunit is a protein of about 268 to 286 amino acid residues whose sequence is well conserved in eukaryotic species [1]. As signature patterns two highly conserved regions in the C-terminal section of the alpha subunit were selected.

Consensus pattern: V-H-[FY](2)-E-D-G-N-V

Consensus pattern: F-K-[AE]-L-R-R-x-L-P-

[1] Cooper J.A., Caldwell J.E., Gattermeir D.J., Torres M.A., Amatruda J.F., Casella J.F. Cell Motil. Cytoskeleton 18:204-214(1991).

186. F-box domain

[1] Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ, Cell 1996;86:263-274. [2] Skowyra D, Craig KL, Tyers M, Elledge SJ, Harper JW, Cell 1997;91:209-219.

187. F-protein

Negative factor, (F-Protein) or Nef.

[1] Arold S, Franken P, Strub M-P, Hoh F, Benichou S, Benarous R, Dumas C; Medline: 98035457, The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signalling Structure 1997;5:1361-1372.

Nef protein accelerates virulent progression of AIDS by its interaction with cellular proteins involved in signal transduction and host cell activation. Nef has been shown to bind specifically to a subset of the Src kinase family.

Number of members: 1013

188. (FAD_binding_2)
Fumarate reductase / succinate dehydrogenase FAD-binding site

In bacteria two distinct, membrane-bound, enzyme complexes are responsible for the interconversion of fumarate and succinate (EC 1.3.99.1): fumarate reductase (Frd) is used in anaerobic growth, and succinate dehydrogenase (Sdh) is used in aerobic growth. Both complexes consist of two main components: a membrane-extrinsic component composed of a FAD-binding flavoprotein and an iron-sulfur protein; and an hydrophobic component composed of a membrane anchor protein and/or a cytochrome B.

In eukaryotes mitochondrial succinate dehydrogenase (ubiquinone) (EC 1.3.5.1) is an enzyme composed of two subunits: a FAD flavoprotein and and iron-sulfur protein.

The flavoprotein subunit is a protein of about 60 to 70 Kd to which FAD is covalently bound to a histidine residue which is located in the N-terminal section of the protein [1]. The sequence around that histidine is well conserved in Frd and Sdh from various bacterial and eukaryotic species [2] and can be used as a signature pattern.

Consensus patternR-[ST]-H-[ST]-x(2)-A-x-G-G [H is the FAD binding site] Sequences known to belong to this class detected by the pattern ALL.

[1] Blaut M., Whittaker K., Valdovinos A., Ackrell B.A., Gunsalus R.P., Cecchini G. J. Biol. Chem. 264:13599-13604(1989).
[2] Birch-Machin M.A., Farnsworth L., Ackrell B.A., Cochran B., Jackson S., Bindoff L.A., Aitken A., Diamond A.G., Turnbull D.M. J. Biol. Chem. 267:11553-11558(1992).

189. Fatty acid desaturases signatures (FA_desaturase)
Fatty acid desaturases (EC 1.14.99.-) are enzymes that catalyze the insertion of a double bond at the delta position of fatty acids. There seems to be two distinct families of fatty acid desaturases which do not seem to be evolutionary related. Family 1 is composed of: -

Stearoyl-CoA desaturase (SCD) (EC 1.14.99.5) [1]. SCD is a key regulatory enzyme of unsaturated fatty acid biosynthesis. SCD introduces a cis double bond at the delta(9) position of fatty acyl-CoA's such as palmitoleoyl- and oleoyl-CoA. SCD is a membrane-bound enzyme that is thought to function as a part of a multienzyme complex in the endoplasmic reticulum of vertebrates and fungi. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected, this region is rich in histidine residues and in aromatic residues. Family 2 is composed of: - Plants stearoyl-acyl-carrier-protein desaturase (EC 1.14.99.6) [2], these enzymes catalyze the introduction of a double bond at the delta(9) position of steraoyl-ACP to produce oleoyl-ACP. This enzyme is responsible for the conversion of saturated fatty acids to unsaturated fatty acids in the synthesis of vegetable oils. - Cyanobacteria desA [3] an enzyme that can introduce a second cis double bond at the delta(12) position of fatty acid bound to membranes glycerolipids. DesA is involved in chilling tolerance; the phase transition temperature of lipids of cellular membranes being dependent on the degree of unsaturation of fatty acids of the membrane lipids. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected.

Consensus pattern: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y-

Consensus pattern: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]- [DE]-

[1] Kaestner K.H., Ntambi J.M., Kelly T.J. Jr., Lane M.D. J. Biol. Chem. 264:14755-14761(1989).

[2] Shanklin J., Somerville C.R. Proc. Natl. Acad. Sci. U.S.A. 88:2510-2514(1991).

[3] Wada H., Gombos Z., Murata N. Nature 347:200-203(1990).

190. Fructose-1-6-bisphosphatase active site (FBPase)

Fructose-1,6-bisphosphatase (EC 3.1.3.11) (FBPase) [1], a regulatory enzyme in gluconeogenesis, catalyzes the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate. It is involved in many different metabolic pathways and found in most organisms.Sedoheptulose-1,7-bisphosphatase (EC 3.1.3.37) (SBPase) [2] is an enzyme found plant chloroplast and in photosynthetic bacteria that catalyzes the hydrolysis of sedoheptulose 1,7-bisphosphate to sedoheptulose 7-phosphate, a step in the Calvin's reductive pentose phosphate cycle. It is functionally and structurally related to FBPase. In mammalian FBPase,

a lysine residue has been shown to be involved in the catalytic mechanism [3]. The region around this residue is highly conserved and can be used as a signature pattern for FBPase and SBPase. It must be noted that, in some bacterial FBPase sequences, the active site lysine is replaced by an arginine

Consensus pattern: [AG]-[RK]-L-x(1,2)-[LIV]-[FY]-E-x(2)-P-[LIVM]-[GSA] [K/R is the active site residue]-

[1] Benkovic S.J., DeMaine M.M. Adv. Enzymol. 53:45-82(1982).
[2] Raines C.A., Lloyd J.C., Willingham N.M., Potts S., Dyer T.A. Eur. J. Biochem. 205:1053-1059(1992).
[3] Ke H., Thorpe C.M., Seaton B.A., Lipscomb W.N., Marcus F. J. Mol. Biol. 212:513-539(1989).

191. FGGY family of carbohydrate kinases signatures *

It has been shown [1] that four different type of carbohydrate kinases seem to be evolutionary related. These enzymes are: - L-fucolokinase (EC 2.7.1.51) (gene fucK). - Gluconokinase (EC 2.7.1.12) (gene gntK). - Glycerokinase (EC 2.7.1.30) (gene glpK). - Xylulokinase (EC 2.7.1.17) (gene xylB). - L-xylulose kinase (EC 2.7.1.53) (gene lyxK).These enzymes are proteins of from 480 to 520 amino acid residues. As consensus patterns for this family of kinases two conserved regionswere selected, one in the central section, the other in the C-terminal section.

Consensus pattern: [MFYGS]-x-[PST]-x(2)-K-[LIVMFYW]-x-W-[LIVMF]-x-[DENQTKR]-[ENQH]-

Consensus pattern: [GSA]-x-[LIVMFYW]-x-G-[LIVM]-x(7,8)-[HDENQ]-[LIVMF]-x(2)-[AS]-[STAIVM]-[LIVMFY]-[DEQ]-

[1] Reizer A., Deutscher J., Saier M.H. Jr., Reizer J. Mol. Microbiol. 5:1081-1089(1991).

192. FKBP-type peptidyl-prolyl cis-trans isomerase signatures/profile (FKBP)

FKBP [1,2,3] is the major high-affinity binding protein, in vertebrates, for the immunosuppressive drug FK506. It exhibits peptidyl-prolyl cis-trans isomerase activity (EC 5.2.1.8) (PPIase or rotamase). PPIase is an enzyme that accelerates protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides [4].At least three different forms of FKBP are known in mammalian species: - FKBP-12, which is cytosolic and inhibited by both FK506 and rapamycin. - FKBP-13, which is membrane associated and inhibited by both FK506 and rapamycin. - FKBP-25, which is preferentially inhibited by rapamycin. These forms of FKBP are evolutionary related and show extensive similarities[5,6,7] with the following proteins: - Fungal FKBP. - Mammalian hsp binding immunophilin (HBI) (also called p59). HBI is a protein which binds to hsp90 and contains two FKBP-like domains in its N- terminal section - the first of which seems to be functional. - The C-terminal part of the cell-surface protein mip from Legionella; a protein associated with macrophage infection by an unknown mechanism. - Escherichia coli slyD [8], a protein with a N-terminal FKBP domain followed by an histidine-rich metal-binding domain. - Escherichia coli fkpA. - Escherichia coli fklB (FKBP22). - Escherichia coli slpA. - Bacterial trigger factor (Tig). - Streptomyces hygroscopus and chrysomallus FK506-binding protein. - Chlamydia trachomatis 27 Kd membrane protein. - Neisseria meningitidis strain C114 PPiase. - Probable PPiases from Haemophilus influenzae (HI0754), Methanococcus jannaschii (MJ0278 and MJ0825), Pseudomonas fluorescens and Pseudomonase aeruginosa. Two signature patterns for these proteins were developed. One is based on a conserved region in the N-terminus of FKBP, the other is located in the central section. The profile for FKBP spans the complete domain.

Consensus pattern: [LIVMC]-x-[YF]-x-[GVL]-x(1,2)-[LFT]-x(2)-G-x(3)-[DE]- [STAEQK]-[STAN]-

Consensus pattern: [LIVMFY]-x(2)-[GA]-x(3,4)-[LIVMF]-x(2)-[LIVMFHK]-x(2)-G- x(4)-[LIVMF]-x(3)-[PSGAQ]-x(2)-[AG]-[FY]-G--

[1] Tropschug M., Wachter E., Mayer S., Schoenbrunner E.R., Schmid F.X. Nature 346:674-677(1990).

[2] Stein R.L. Curr. Biol. 1:234-236(1991).

[3] Siekierka J.J., Widerrecht G., Greulich H., Boulton D., Hung S.H.Y., Cryan J., Hodges P.J., Sigal N.H. J. Biol. Chem. 265:21011-21015(1990).

[4] Fischer G., Schmid F.X. Biochemistry 29:2205-2212(1990).

[5] Trandinh C.C., Pao G.M., Saier M.H. Jr. FASEB J. 6:3410-3420(1992).

[6] Galat A. Eur. J. Biochem. 216:689-707(1993).

[7] Hacker J., Fischer G. Mol. Microbiol. 10:445456(1993).

[8] Wuelfing C., Lomardero J., Plueckthun A. J. Biol. Chem. 269:2895-2901(1994).

193. MAPEG family (aka: FLAP/GST2/LTC4S family signature)

The following mammalian proteins are evolutionary related [1]:

- Leukotriene C4 synthase (EC 2.5.1.37) (gene LTC4S), an enzyme that catalyzes the production of LTC4 from LTA4.
- Microsomal glutathione S-transferase II (EC 2.5.1.18) (GST-II) (gene GST2), an enzyme that can also produces LTC4 fron LTA4.
- 5-lipoxygenase activating protein (gene FLAP), a protein that seems to be required for the activation of 5-lipoxygenase.

These are proteins of 150 to 160 residues that contain three transmembrane segments. As a signature pattern, a conserved region between the first and second transmembrane domains was selected.

Consensus patternc: G-x(3)-F-E-R-V-[FY]-x-A-[NQ]-x-N-C

[1] Jakobsson P.-J., Mancini J.A., Ford-Hutchinson A.W. J. Biol. Chem. 271:22203-22210(1996).

194. FMN-dependent alpha-hydroxy acid dehydrogenases active site (FMN_dh)

A number of oxidoreductases that act on alpha-hydroxy acids and which are FMN-containing flavoproteins have been shown [1,2,3] to be structurally related; these enzymes are: - Lactate dehydrogenase (EC 1.1.2.3), which consists of a dehydrogenase domain and a heme-binding domain called cytochrome b2 and which catalyzes the conversion of lactate into pyruvate. - Glycolate oxidase (EC 1.1.3.15) ((S)-2-hydroxy-acid oxidase), a peroxisomal enzyme that catalyzes the conversion of glycolate and oxygen to glyoxylate and hydrogen peroxide. - Long chain alpha-hydroxy acid oxidase from rat (EC 1.1.3.15), a peroxisomal enzyme. - Lactate 2-monooxygenase (EC 1.13.12.4) (lactate oxidase) from Mycobacterium smegmatis, which catalyzes the conversion of lactate and oxygen to acetate, carbon dioxide and water. - (S)-mandelate dehydrogenase from Pseudomonas putida (gene mdlB), which catalyzes the

reduction of (S)-mandelate to benzoylformate. The first step in the reaction mechanism of these enzymes is the abstraction of the proton from the alpha-carbon of the substrate producing a carbanion which can subsequently attach to the N5 atom of FMN. A conserved histidine has been shown [4] to be involved in the removal of the proton. The region around this active site residue is highly conserved and contains an arginine residue which is involved in substrate binding.

Consensus pattern: S-N-H-G-[AG]-R-Q [H is the active site residue] [R is a substrate-binding residue]-

[1] Giegel D.A., Williams C.H. Jr., Massey V. J. Biol. Chem. 265:6626-6632(1990).
[2] Tsou A.Y., Ransom S.C., Gerlt J.A., Buechter D.D., Babbitt P.C., Kenyon G.L. Biochemistry 29:9856-9862(1990).
[3] Le K.H.D., Lederer F. J. Biol. Chem. 266:20877-20880(1991).
[4] Lindqvist Y., Branden C.-I. J. Biol. Chem. 264:3624-3628(1989).

195. Flavin-binding monooxygenase-like (FMO-like)

This family includes FMO proteins, cyclohexanone monooxygenase

196. (FPGS)

Folylpolyglutamate synthase signatures (aka Mur_ligase)

Folylpolyglutamate synthase (EC 6.3.2.17) (FPGS) [1] is the enzyme of folate metabolism that catalyzes ATP-dependent addition of glutamate moieties to tetrahydrofolate.

Its sequence is moderately conserved between prokaryotes (gene folC) and eukaryotes. We developed two signature patterns based on the conserved regions which are rich in glycine residues and could play a role in the catalytical activity and/or in substrate binding.

Consensus pattern [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)- [LIVM](2)-x(3)-[GSK] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern[LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL.

[1] Shane B., Garrow T., Brenner A., Chen L., Choi Y.J., Hsu J.C., Stover P. Adv. Exp. Med. Biol. 338:629-634(1993).

197. FYVE zinc finger

The FYVE zinc finger is named after four proteins that it has been found in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger has been shown to bind two Zn++ ions [1]. The FYVE finger has eight potential zinc coordinating cysteine positions. Many members of this family also include two histidines in a motif R+HHC+XCG, where + represents a charged residue and X any residue. Members were included which do not conserve these histidine residues but are clearly related.

[1] Stenmark H, Aasland R, Toh BH, D'Arrigo A, J Biol Chem 1996;271:24048-24054. [2] Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Stenmark H, Aasland R, Nature 1998;394:432-433.

198. F_actin_cap_B

F-actin capping protein beta subunit signature

The F-actin capping protein binds in a calcium-independent manner to the fast growing ends of actin filaments (barbed end) thereby blocking the exchange of subunits at these ends. Unlike gelsolin and severin this protein does not sever actin filaments. The F-actin capping protein is a heterodimer composed of two unrelated subunits: alpha and beta.

The beta subunit is a protein of about 280 amino acid residues whose sequence is well conserved in eukaryotic species [1]. As a signature pattern a conserved hexapeptide in the N-terminal section of the beta subunit was selected.

Consensus pattern: C-D-Y-N-R-D Sequences known to belong to this class detected by the pattern ALL.

[1] Amatruda J.F., Cannon J.F., Tatchell K., Hug C., Cooper J.A. Nature 344:352-354(1990).

199. Isopenicillin N synthetase signatures (Fe_Asc_oxidored)

Isopenicillin N synthetase (IPNS) [1,2] is a key enzyme in the biosynthesis of penicillin and cephalosporin. In the presence of oxygen, it removes iron and ascorbate, four hydrogen atoms from L-(alpha-aminoadipyl)-L-cysteinyl-d-valine to form the azetidinone and thiazolidine rings of isopenicillin. IPNS is an enzyme of about 330 amino-acid residues. Two cysteines are conserved in fungal and bacterial IPNS sequences; these may be involved in iron-binding and/or substrate-binding. Cephalosporium acremonium DAOCS/DACS [3] is a bifunctional enzyme involved in cephalosporin biosynthesis. The DAOCS domain, which is structurally related to IPNS, catalyzes the step from penicillin N to deacetoxy-cephalosporin C - used as a substrate by DACS to form deacetylcephalosporin C. Streptomycesclavuligerus possesses a monofunctional DAOCS enzyme (gene cefE) [4] also related to IPNS. Two signature patterns for these enzymes were derived, centered around the conserved cysteine residues.

Consensus pattern: [RK]-x-[STA]-x(2)-S-x-C-Y-[SL]-

Consensus pattern: [LIVM](2)-x-C-G-[STA]-x(2)-[STAG]-x(2)-T-x-[DNG]-

[1] Martin J.F. Trends Biotechnol. 5:306-308(1987).

[2] Chen G., Shiffman D., Mevarech M., Aharonowitz Y. Trends Biotechnol. 8:105-111(1990).

[3] Samson S.M., Dotzlaf J.E., Slisz M.L., Becker G.W., van Frank R.M., Veal L.E., Yeh W.K., Miller J.R., Queener S.W., Ingolia T.D. Bio/Technology 5:1207-1214(1987).

[4] Kovacevic S., Weigel B.J., Tobin M.B., Ingolia T.D., Miller J.R. J. Bacteriol. 171:754-760(1989).

200. Fibrillarin signature

Fibrillarin [1] is a component of a nucleolar small nuclear ribonucleoprotein(SnRNP) particle thought to participate in the first step of the processing of pre-rRNA. In mammals, fibrillarin is associated with the U3, U8 and U13small nuclear RNAs [2]. Fibrillarin is an extremely well conserved protein of about 320 amino acid residues. Structurally it consists of three

different domains: - An N-terminal domain of about 80 amino acids which is very rich in glycine and contains a number of dimethylated arginine residues (DMA). - A central domain of about 90 residues which resembles that of RNA-binding proteins and contains an octameric sequence similar to the RNP-2 consensus found in such proteins. - A C-terminal alpha-helical domain. A protein evolutionary related to fibrillarin has been found [3] in archaebacteria such as Methanococcus vannielii or voltae. This protein (geneflpA) is involved in pre-rRNA processing. It lacks the Gly/Arg-rich N-terminal domain. As a signature pattern, a region was selected that starts with and encompases theRNP-2 like octapeptide sequence.

Consensus pattern: [GST]-[LIVMAP]-V-Y-A-[IV]-E-[FY]-[SA]-x-R-x(2)-R-[DE] -

[1] Aris J.P., Blobel G. Proc. Natl. Acad. Sci. U.S.A. 88:931-935(1991).
[2] Bandziulis R.J., Swanson M.S., Dreyfuss G. Genes Dev. 3:431-437(1989).
[3] Agha-Amiri K. J. Bacteriol. 176:2124-2127(1994).

201. Filamin/ABP280 repeat

[1] Fucini P, Renner C, Herberhold C, Noegel AA, Holak TA, Nat Struct Biol 1997;4:223-230.

202. Fucosyl transferase

This family of Fucosyltransferases are the enzymes transferring fucose from GDP-Fucose to GlcNAc in an alpha1,3 linkage [1].

[1] Breton C, Oriol R, Imberty A; Glycobiology 1998;8:87-94.

203. 2Fe-2S ferredoxins, iron-sulfur binding region signature (fer2A)

Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One of these subgroups are the 2Fe-2S ferredoxins, which are proteins or domains of around one hundred amino acid residues that bind a single 2Fe-2S iron-sulfur

cluster. The proteins that are known [2] to belong to this family are listed below. - Ferredoxin from photosynthetic organisms; namely plants and algae where it is located in the chloroplast or cyanelle; and cyanobacteria. - Ferredoxin from archaebacteria of the Halobacterium genus. - Ferredoxin IV (gene pftA) and V (gene fdxD) from Rhodobacter capsulatus. - Ferredoxin in the toluene degradation operon (gene xylT) and naphthalene degradation operon (gene nahT) of Pseudomonas putida. - Hypothetical Escherichia coli protein yfaE. - The N-terminal domain of the bifunctional ferredoxin/ferredoxin reductase electron transfer component of the benzoate 1,2-dioxygenase complex (gene benC) from Acinetobacter calcoaceticus, the toluene 4-monooxygenase complex (gene tmoF), the toluate 1,2-dioxygenase system (gene xylZ), and the xylene monooxygenase system (gene xylA) from Pseudomonas. - The N-terminal domain of phenol hydroxylase protein p5 (gene dmpP) from Pseudomonas Putida. - The N-terminal domain of methane monooxygenase component C (gene mmoC) from Methylococcus capsulatus . - The C-terminal domain of the vanillate degradation pathway protein vanB in a Pseudomonas species. - The N-terminal domain of bacterial fumarate reductase iron-sulfur protein (gene frdB). - The N-terminal domain of CDP-6-deoxy-3,4-glucoseen reductase (gene ascD) from Yersinia pseudotuberculosis. - The central domain of eukaryotic succinate dehydrogenase (ubiquinone) iron- sulfur protein. - The N-terminal domain of eukaryotic xanthine dehydrogenase. - The N-terminal domain of eukaryotic aldehyde oxidase. In the 2Fe-2S ferredoxins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region.

Consensus pattern: C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C [The three C's are 2Fe-2S ligands]-

[1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).[2] Harayama S., Polissi A., Rekik M. FEBS Lett. 285:85-88(1991).

Adrenodoxin family, iron-sulfur binding region signature (fer2B)

Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One family of ferredoxins groups together the following proteins that all bind a single 2Fe-2S iron-sulfur cluster: - Adrenodoxin (ADX) (adrenal ferredoxin), a vertebrate

mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450scc, which is involved in cholesterol side chain cleavage. - Putidaredoxin (PTX), a Pseudomonas putida protein which transfers electrons from putidaredoxin reductase to cytochrome P450-cam, which is involved in the oxidation of camphor. - Terpredoxin [2], a Pseudomonas protein which transfers electrons from terpredoxin reductase to cytochrome P450-terp, which is involved in the oxidation of alpha-terpineol. - Rhodocoxin [3], a Rhodococcus protein which transfers electrons from rhodocoxin reductase to cytochrome CYP116 (thcB), which is involved in the degradation of thiocarbamate herbicides. - Escherichia coli ferredoxin (gene fdx) [4] whose exact function is not yet known. - Rhodobacter capsulatus ferredoxin VI [5], which may transfer electrons to a yet uncharacterized oxygenase. - Caulobacter crescentus ferredoxin (gene fdxB) [6].In these proteins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region.

Consensus pattern: C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR] [The three C's are 2Fe-2S ligands]-

[1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).
[2] Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S., Carmona C., Witney F., Lorence M.C. J. Biol. Chem. 267:14193-14203(1992).
[3] Nagy I., Schoofs G., Compernolle F., Proost P., Vanderleyden J., De Mot R. J. Bacteriol. 177:676-687(1995).
[4] Ta D.T., Vickery L.E. J. Biol. Chem. 267:11120-11125(1992).
[5] Naud I., Vincon M., Garin J., Gaillard J., Forest E., Jouanneau Y. Eur. J. Biochem. 222:933-939(1994).
[6] Amemiya K EMBL/Genbank: X51607.

204. 4Fe-4S ferredoxins, iron-sulfur binding region signature (fer4)
Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron-sulfur cluster(s). One of these subgroups are the 4Fe-4S ferredoxins, which are found in bacteria and which are thus often referred as

'bacterial-type' ferredoxins. The structure of these proteins [2] consists of the duplication of a domain of twenty six amino acid residues; each of these domains contains four cysteine residues that bind to a 4Fe-4S center. A number of proteins have been found [3] that include one or more 4Fe-4Sbinding domains similar to those of bacterial-type ferredoxins. These proteins are listed below (references are only provided for recently determined sequences). - The iron-sulfur proteins of the succinate dehydrogenase and the fumarate reductase complexes (EC 1.3.99.1). These enzyme complexes, which are components of the tricarboxylic acid cycle, each contain three subunits: a flavoprotein, an iron-sulfur protein, and a b-type cytochrome. The iron- sulfur proteins contain three different iron-sulfur centers: a 2Fe-2S, a 3Fe-3S and a 4Fe-4S. - Escherichia coli anaerobic glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) This enzyme is composed of three subunits: A, B, and C. The C subunit seems to be an iron-sulfur protein with two ferredoxin-like domains in the N-terminal part of the protein. - Escherichia coli anaerobic dimethyl sulfoxide reductase. The B subunit of this enzyme (gene dmsB) is an iron-sulfur protein with four 4Fe-4S ferredoxin-like domains. - Escherichia coli formate hydrogenlyase. Two of the subunits of this oligomeric complex (genes hycB and hycF) seem to be iron-sulfur proteins that each contain two 4Fe-4S ferredoxin-like domains. - Methanobacterium formicicum formate dehydrogenase (EC 1.2.1.2). This enzyme is used by the archaebacteria to grow on formate. The beta chain of this dimeric enzyme probably binds two 4Fe-4S centers. - Escherichia coli formate dehydrogenases N and O (EC 1.2.1.2). The beta chain of these two enzymes (genes fdnH and fdoH) are iron-sulfur proteins with four 4Fe-4S ferredoxin-like domains. - Desulfovibrio periplasmic [Fe] hydrogenase (EC 1.18.99.1). The large chain of this dimeric enzyme binds three 4Fe-4S centers, two of which are located in the ferredoxin-like N-terminal region of the protein. - Methanobacterium thermoautrophicum methyl viologen-reducing hydrogenase subunit mvhB, which contains six tandemly repeated ferredoxin-like domains and which probably binds twelve 4Fe-4S centers. - Salmonella typhimurium anaerobic sulfite reductase (EC 1.8.1.-) [4]. Two of the subunits of this enzyme (genes asrA and asrC) seem to both bind two 4Fe-4S centers. - A Ferredoxin-like protein (gene fixX) from the nitrogen-fixation genes locus of various Rhizobium species, and one from the Nif-region of Azotobacter species. - The 9 Kd polypeptide of chloroplast photosystem I [5] (gene psaC). This protein contains two low potential 4Fe-4S centers, referred as the A and B centers. - The chloroplast frxB protein which is predicted to carry two 4Fe-4S centers. - An ferredoxin from a primitive eukaryote, the enteric amoeba Entamobea histolytica. - Escherichia coli hypothetical protein yjjW, a protein with a N-terminal region belonging to the radical activating enzymes family

(see <PDOC00834>) and two potential 4Fe-4S centers.The pattern of cysteine residues in the iron-sulfur region is sufficient todetect this class of 4Fe-4S binding proteins.

Consensus pattern: C-x(2)-C-x(2)-C-x(3)-C-[PEG] [The four C's are 4Fe-4S ligands]-

[1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).
[2] Otaka E., Ooi T. J. Mol. Evol. 26:257-267(1987).
[3] Beinert H. FASEB J. 4:2483-2492(1990).
[4] Huang C.J., Barrett E.L. J. Bacteriol. 173:1544-1553(1991).
[5] Knaff D.B. Trends Biochem. Sci. 13:460-461(1988).

205. NifH/frxC family signatures (fer4_NifH)

Nitrogenase (EC 1.18.6.1) [1] is the enzyme system responsible for biological nitrogen fixation. Nitrogenase is an oligomeric complex which consists of two components: component 1 which contains the active site for the reduction of nitrogen to ammonia and component 2 (also called the iron protein).Component 2 is a homodimer of a protein (gene nifH) which binds a single 4Fe-4S iron sulfur cluster [2]. In the nitrogen fixation process nifH is first reduced by a protein such as ferredoxin; the reduced protein then transfers electrons to component 1 with the concomitant consumption of ATP.A number of proteins are known to be evolutionary related to nifH. These proteins are: - Chloroplast encoded frxC (or chlL) protein [3]. FrxC is encoded on the chloroplast genome of some plant species, its exact function is not known, but it could act as an electron carrier in the conversion of protochlorophyllide to chlorophyllide. - Rhodobacter capsulatus proteins bchL and bchX [4]. These proteins are also likely to play a role in chlorophyll synthesis. There are a number of conserved regions in the sequence of these proteins: in the N-terminal section there is an ATP-binding site motif 'A' (P-loop) and in the central section there are two conserved cysteines which have been shown, in nifH, to be the ligands of the 4Fe-4S cluster. Two signatures patterns that correspond to the regions around these cysteines were developed.

Consensus pattern: E-x-G-G-P-x(2)-[GA]-x-G-C-[AG]-G [C binds the iron-sulfur center]-
Consensus pattern: D-x-L-G-D-V-V-C-G-G-F-[AG]-x-P [C binds the iron-sulfur center]-

[1] Pau R.N. Trends Biochem. Sci. 14:183-186(1989).

[2] Georgiadis M.M., Komiya H., Chakrabarti P., Woo D., Kornuc J.J., Rees D.C. Science 257:1653-1659(1992).

[3] Fujita Y., Takahashi Y., Kohchi T., Ozeki H., Ohyama K., Matsubara H. Plant Mol. Biol. 13:551-561(1989).

[4] Burke D.H., Alberti M., Hearst J.E. J. Bacteriol. 175:2407-2413(1993).

206. Ferritin iron-binding regions signatures

Ferritin [1,2] is one of the major non-heme iron storage proteins. It consists of a mineral core of hydrated ferric oxide, and a multi-subunit protein shell which englobes the former and assures its solubility in an aqueous environment. In animals the protein is mainly cytoplasmic and there are generally two or more genes that encodes for closely related subunits (in mammals there are two subunits which are known as H(eavy) and L(ight)). In plants ferritin is found in the chloroplast [3].There are a number of well conserved region in the sequence of ferritins. Two of these regions to develop signature patterns were selected. The first pattern is located in the central part of the sequence of ferritin and it contains three conserved glutamate which are thought to be involved in the binding of iron. The second pattern is located in the C-terminal section, it corresponds to a region which forms a hydrophilic channel through which small molecules and ions can gain access to the central cavity of the molecule; this pattern also includes conserved acidic residues which are potential metal-binding sites.

Consensus pattern: E-x-[KR]-E-x(2)-E-[KR]-[LF]-[LIVMA]-x(2)-Q-N-x-R-x-G-R [The 3 E's are potential iron ligands]-

Consensus pattern: D-x(2)-[LIVMF]-[STAC]-[DH]-F-[LI]-[EN]-x(2)-[FY]-L-x(6)- [LIVM]-[KN] [The second D and the E are potential iron ligands]-

[1] Crichton R.R., Charloteaux-Wauters M. Eur. J. Biochem. 164:485-506(1987).

[2] Theil E.C. Annu. Rev. Biochem. 56:289-315(1987).

[3] Ragland M., Briat J.-F., Gagnon J., Laulhere J.-P., Massenet O., Theil E.C. J. Biol. Chem. 265:18339-18344(1990).

207. Intermediate filaments signature (filament)

Intermediate filaments (IF) [1,2,3] are proteins which are primordial components of the cytoskeleton and the nuclear envelope. They generally form filamentous structures 8 to 14 nm wide. IF proteins are members of a very large multigene family of proteins which has been subdivided in five major subgroups: - Type I: Acidic cytokeratins. - Type II: Basic cytokeratins. - Type III: Vimentin, desmin, glial fibrillary acidic protein (GFAP), peripherin, and plasticin. - Type IV: Neurofilaments L, H and M, alpha-internexin and nestin. - Type V: Nuclear lamins A, B1, B2 and C. All IF proteins are structurally similar in that they consist of: a central rod domain comprising some 300 to 350 residues which is arranged in coiled-coiled alpha-helices, with at least two short characteristic interruptions; a N-terminal non-helical domain (head) of variable length; and a C-terminal domain (tail) which is also non-helical, and which shows extreme length variation between different IF proteins. While IF proteins are evolutionary and structurally related, they have limited sequence homologies except in several regions of the rod domain. A conserved region at the C-terminal extremity of the rod domain was used as a sequence pattern for this class of proteins.

Consensus pattern: [IV]-x-[TACI]-Y-[RKH]-x-[LM]-L-[DE]-

[1] Quinlan R., Hutchison C., Lane B. Protein Prof. 2:801-952(1995).
[2] Steiner P.M., Roop D.R. Annu. Rev. Biochem. 57:593-625(1988).
[3] Stewart M. Curr. Opin. Cell Biol. 2:91-100(1990).

208. Flavodoxin signature

Flavodoxins [1,E1] are electron-transfer proteins that function in various electron transport systems. Flavodoxins bind one FMN molecule, which serves as a redox-active prosthetic group. Flavodoxins are functionally interchangeable with ferredoxins. They have been isolated from prokaryotes, cyanobacteria, and some eukaryotic algae. The signature pattern for these proteins is derived from a conserved region in their N-terminal section, this region is involved in the binding of the FMN phosphate group.

Consensus pattern: [LIV]-[LIVFY]-[FY]-x-[ST]-x(2)-[AGC]-x-T-x(3)-A-x(2)-[LIV]-

[1] Wakabayashi S., Kimura K., Matsubara H., Rogers L.J. Biochem. J. 263:981-984(1989).

209. Growth factor and cytokines receptors family signatures (fn3)

A number of receptors for lymphokines, hematopoeitic growth factors and growth hormone-related molecules have been found [1 to 5] to share a common binding domain. Receptors known to belong to this family are: - Cytokine receptor common beta chain. This chain is common to the IL-3, IL-5 and GM-CSF receptors. - Cytokine receptor common gamma chain. This chain is common to the IL-2, IL-4, IL-7 and IL-13 receptors. - Ciliary neurotrophic factor receptor (CNTFR). - Erythropoietin receptor (EPOR). - Granulocyte colony-stimulating factor receptor (G-CSFR). - Granulocyte-macrophage colony-stimulating factor receptor alpha chain (GM- CSFR). - Interleukin-2 receptor beta chain (IL2R-beta). - Interleukin-3 receptor alpha chain (IL3R). - Interleukin-4 receptor alpha chain (IL4R). - Interleukin-5 receptor alpha chain (IL5R). - Interleukin-6 receptor (IL6R). - Interleukin-7 receptor alpha chain (IL7R). - Interleukin-9 receptor (IL9R). - Growth hormone receptor (GRHR). - Prolactin receptor (PRLR). - Thrombopoeitin receptor (TPOR).The conserved region constitutes all or part of the extracellular ligand-binding region and is about 200 amino acid residues long. In the N-terminal of this domain there are two pairs of cysteines known, in the growth hormone receptor, to be involved in disulfide bonds. +---------------------------------------xxxxxxx--------------------------+ | C C C C Extracellular XXXXXXX Cytoplasmic | +-|-|-------|--|------------------------xxxxxxx--------------------------+ | | | | Transmembrane +-+ +--+ Two patterns to detect this family of receptors were used. The first one is derived from the first N-terminal disulfide loop, the second is a tryptophan-rich pattern located at the C-terminal extremity of the extracellular region.

Consensus pattern: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W [The two C's are linked by a disulfide bond]-

Consensus pattern: [STGL]-x-W-[SG]-x-W-S-

[1] Bazan J.F. Biochem. Biophys. Res. Commun. 164:788-795(1989).

[2] Bazan J.F. Proc. Natl. Acad. Sci. U.S.A. 87:6934-6938(1990).

[3] Cosman D., Lyman S.D., Idzerda R.L., Beckmann M.P., Park L.S., Goodwin R.G., March C.J. Trends Biochem. Sci. 15:265-270(1990).

[4] d'Andrea A.D., Fasman G.D., Lodish H.F. Cell 58:1023-1024(1989).

[5] d'Andrea A.D., Fasman G.D., Lodish H.F. Curr. Opin. Cell Biol. 2:648-651(1990).

210. Phosphoribosylglycinamide formyltransferase active site (formyl_transf)
Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) [1] catalyzes the third step in de novo purine biosynthesis, the transfer of a formyl group to 5'-phosphoribosylglycinamide. In higher eukaryotes, GART is part of a multifunctional enzyme polypeptide that catalyzes three of the steps of purine biosynthesis. In bacteria, plants and yeast, GART is a monofunctional protein of about 200 amino-acid residues. In the Escherichia coli enzyme, an aspartic acid residue has been shown to be involved in the catalytic mechanism. The region around this active site residue is well conserved in GART from prokaryotic and eukaryotic sources and can be used as a signature pattern. Mammalian formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) [2] is a cytosolicenzyme responsible for the NADP-dependent decarboxylative reduction of 10-formyltetrahydrofolate into tetrahydrofolate. It is a protein of about 900 amino acids consisting of three domains; the N-terminal domain (200 residues) is structurally related to GARTs.Escherichia coli methionyl-tRNA formyltransferase (EC 2.1.2.9) (gene fmt) [3]is the enzyme responsible for modifying the free amino group of the aminoacylmoiety of methionyl- A(fMet). The central part of fmt seems to be evolutionary related to GART's active site region.

Consensus pattern: G-x-[STM]-[IVT]-x-[FYWVQ]-[VMAT]-x-[DEVM]-x-[LIVMY]-D-x-G- x(2)-[LIVT]-x(6)-[LIVM] [D is the active site residue] -

[1] Inglese J., Smith J.M., Benkovic S.J. Biochemistry 29:6678-6687(1990).
[2] Cook R.J., Lloyd R.S., Wagner C. J. Biol. Chem. 266:4965-4973(1991).
[3] Guillon J.-M., Mechulam Y., Schmitter J.-M., Blanquet S., Fayat G. J. Bacteriol. 174:4294-4301(1992).

211. G10 protein signatures
A Xenopus protein known as G10 [1] has been found to be highly conserved in a wide range of eukaryotic species. The function of G10 is still unknown. G10 is a protein of about 17 to 18 Kd (143 to 157 residues) which is hydrophilic and whose C-terminal half is rich in cysteines and could be involved in metal-binding. As signature patterns, two of these cysteine-rich segments were selected.

Consensus pattern: L-C-C-x-[KR]-C-x(4)-[DE]-x-N-x(4)-C-x-C-R-V-P-

Consensus pattern: C-x-H-C-G-C-[KRH]-G-C-[SA]-

[1] McGrew L.L., Dworkin-Rastl E., Dworkin M.B., Richter J.D. Genes Dev. 3:803-815(1989).

212. G-protein alpha subunit

G proteins couple receptors of extracellular signals to intracellular signaling pathways. The G protein alpha subunit binds guanyl nucleotide and is a weak GTPase.

Number of members: 195

[1] Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR, Science 1994;265:1405-1412.

[2] How G proteins work: a continuing story. Coleman DE, Sprang SR, Trends Biochem Sci 1996;21:41-44.

213. Glucose-6-phosphate dehydrogenase active site (G6PD)

Glucose-6-phosphate dehydrogenase (EC 1.1.1.49) (G6PD) [1] catalyzes the first step in the pentose pathway, the reduction of glucose-6-phosphate to gluconolactone 6-phosphate. A lysine residue has been identified as are active nucleophile associated with the activity of the enzyme. The sequence around this lysine is totally conserved from bacterial to mammalian G6PD's and can be used as a signature pattern

Consensus pattern: D-H-Y-L-G-K-[EQK] [K is the active site residue]-

[1] Jeffery J., Persson B., Wood I., Bergman T., Jeffery R., Joernvall H. Eur. J. Biochem. 212:41-49(1993).

214. GATA-type zinc finger domain

The GATA family of transcription factors are proteins that bind to DNA sites with the consensus sequence (A/T)GATA(A/G), found within the regulatory region of a number of

genes. Proteins currently known to belong to this family are: - GATA-1 [1] (also known as Eryf1, GF-1 or NF-E1), which binds to the GATA region of globin genes and other genes expressed in erythroid cells. It is a transcriptional activator which probably serves as a general 'switch' factor for erythroid development. - GATA-2 [2], a transcriptional activator which regulates endothelin-1 gene expression in endothelial cells. - GATA-3 [3], a transcriptional activator which binds to the enhancer of the T-cell receptor alpha and delta genes. - GATA-4 [4], a transcriptional activator expressed in endodermally derived tissues and heart. - Drosophila protein pannier (or DGATAa) (gene pnr) which acts as a repressor of the achaete-scute complex (as-c). - Bombyx mori BCFI [5], which regulates the expression of chorion genes. - Caenorhabditis elegans elt-1 and elt-2, transcriptional activators of genes containing the GATA region, including vitellogenin genes [6]. - Ustilago maydis urbs1 [7], a protein involved in the repression of the biosynthesis of siderophores. - Fission yeast protein GAF2.All these transcription factors contain a pair of highly similar 'zinc finger' type domains with the consensus sequence C-x2-C-x17-C-x2-C.Some other proteins contain a single zinc finger motif highly related to those of the GATA transcription factors. These proteins are: - Drosophila box A-binding factor (ABF) (also known as protein serpent (gene srp)) which may function as a transcriptional activator protein and may play a key role in the organogenesis of the fat body. - Emericella nidulans areA [8], a transcriptional activator which mediates nitrogen metabolite repression. - Neurospora crassa nit-2 [9], a transcriptional activator which turns on the expression of genes coding for enzymes required for the use of a variety of secondary nitrogen sources, during conditions of nitrogen limitation. - Neurospora crassa white collar proteins 1 and 2 (WC-1 and WC-2), which control expression of light-regulated genes. - Saccharomyces cerevisiae DAL81 (or UGA43), a negative nitrogen regulatory protein. - Saccharomyces cerevisiae GLN3, a positive nitrogen regulatory protein. - Saccharomyces cerevisiae GAT1. - Saccharomyces cerevisiae GZF3.

Consensus pattern: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)- C-N-[AS]-C [The four C's are zinc ligands]

[1] Trainor C.D., Evans T., Felsenfeld G., Boguski M.S. Nature 343:92-96(1990).
[2] Lee M.E., Temizer D.T., Clifford J.A., Quertermous T. J. Biol. Chem. 266:16188-16192(1991).
[3] Ho I.-C., Vorhees P., Marin N., Oakley B.K., Tsai S.-F., Orkin S.H., Leiden J.M. EMBO J. 10:1187-1192(1991).

[4] Spieth J., Shim Y.H., Lea K., Conrad R., Blumenthal T. Mol. Cell. Biol. 11:4651-4659(1991).

[5] Drevet J.R., Skeiky Y.A., Iatrou K. J. Biol. Chem. 269:10660-10667(1994).

[6] Hawkins M.G., McGhee J.D. J. Biol. Chem. 270:14666-14671(1995).

[7] Voisard C.P.O., Wang J., Xu P., Leong S.A., McEvoy J.L. Mol. Cell. Biol. 13:7091-7100(1993).

[8] Arst H.N. Jr., Kudla B., Martinez-Rossi N.M., Caddick M.X., Sibley S., Davies R.W. Trends Genet. 5:291-291(1989).

[9] Fu Y.-H., Marzluf G.A. Mol. Cell. Biol. 10:1056-1065(1990).

215. Glutamine amidotransferases class-I active site (GATase)

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known asglutamine amidotransferase (GATase) (EC 2.4.2.-) [1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I(also known as trpG-type) and class-II (also known as purF-type). Class-I GATase domains have been found in the following enzymes: - The second component of anthranilate synthase (AS) (EC 4.1.3.27) [4]. AS catalyzes the biosynthesis of anthranilate from chorismate and glutamine. AS is generally a dimeric enzyme: the first component can synthesize anthranilate using ammonia rather than glutamine, whereas component II provides the GATase activity. In some bacteria and in fungi the GATase component of AS is part of a multifunctional protein that also catalyzes other steps of the biosynthesis of tryptophan. - The second component of 4-amino-4-deoxychorismate (ADC) synthase (EC 4.1.3. -), a dimeric prokaryotic enzyme that function in the pathway that catalyzes the biosynthesis of para-aminobenzoate (PABA) from chorismate and glutamine. The second component (gene pabA) provides the GATase activity [4]. - CTP synthase (EC 6.3.4.2). CTP synthase catalyzes the final reaction in the biosynthesis of pyrimidine, the ATP-dependent formation of CTP from UTP and glutamine. CTP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the C-terminal section [2]. - GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2). GMP synthase catalyzes the ATP-dependent formation of GMP from xanthosine 5'-phosphate and glutamine. GMP synthase is a single chain enzyme that contains two distinct

domains; the GATase domain is in the N-terminal section [5]. - Glutamine-dependent carbamoyl-phosphate synthase (EC 6.3.5.5) (GD-CPSase); an enzyme involved in both arginine and pyrimidine biosynthesis and which catalyzes the ATP-dependent formation of carbamoyl phosphate from glutamine and carbon dioxide. In bacteria GD-CPSase is composed of two subunits: the large chain (gene carB) provides the CPSase activity, while the small chain (gene carA) provides the GATase activity. In yeast the enzyme involved in arginine biosynthesis is also composed of two subunits: CPA1 (GATase), and CPA2 (CPSase). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme (called URA2 in yeast, rudimentary in Drosophila, and CAD in mammals). The GATase domain is located at the N-terminal extremity of this polyprotein [6]. - Phosphoribosylformylglycinamidine synthase II (EC 6.3.5.3), an enzyme that catalyzes the fourth step in the de novo biosynthesis of purines. In some species of bacteria, FGAM synthase II is composed of two subunits: a small chain (gene purQ) which provides the GATase activity and a large chain (gene purL) which provides the aminator activity. - The histidine amidotransferase hisH, an enzyme that catalyzes the fifth step in the biosynthesis of histidine in prokaryotes.In the second component of AS a cysteine has been shown [7] to be essentialfor the amidotransferase activity. The sequence around this residue is well conserved in all the above GATase domains and can be used as a signature pattern for class-I GATase.-

Consensus pattern: [PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]- x-[LIVMFA] [C is the active site residue]-

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).

[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).

[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).

[4] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).

[5] Zalkin H., Argos P., Narayana S.V.L., Tiedeman A.A., Smith J.M. J. Biol. Chem. 260:3350-3354(1985).

[6] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[7] Tso J.Y., Hermodson M.A., Zalkin H. J. Biol. Chem. 255:1451-1457(1980).

216. Glutamine amidotransferases class-II active site (GATase_2)

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known as glutamine amidotransferase (GATase) (EC 2.4.2.-) [1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I(also known as trpG-type) and class-II (also known as purF-type). Class-II GATase domains have been found in the following enzymes: - Amido phosphoribosyltransferase (glutamine phosphoribosylpyrophosphate amidotransferase) (EC 2.4.2.14). An enzyme which catalyzes the first step in purine biosynthesis, the transfer of the ammonia group of glutamine to PRPP to form 5-phosphoribosylamine (gene purF in bacteria, ADE4 in yeast). - Glucosamine--fructose-6-phosphate aminotransferase (EC 2.6.1.16). This enzyme catalyzes a key reaction in amino sugar synthesis, the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine (gene glmS in Escherichia coli, nodM in Rhizobium, GFA1 in yeast) - Asparagine synthetase (glutamine-hydrolyzing) (EC 6.3.5.4). This enzyme is responsible for the synthesis of asparagine from aspartate and glutamine. A cysteine is present at the N-terminal extremity of the mature form of all these enzymes. The cysteine has been shown, in amido phosphoribosyltransferase [4] and in asparagine synthetase [5] to be important for the catalytic mechanism.

Consensus pattern: <x(0,11)-C-[GS]-[IV]-[LIVMFYW]-[AG] [C is the active site residue]-

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).

[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).

[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).

[4] van Heeke G., Schuster M. J. Biol. Chem. 264:5503-5509(1989).

[5] Vollmer S.J., Switzer R.L., Hermodson M.A., Bower S.G., Zalkin H. J. Biol. Chem. 258:10582-10585(1983).

217. GDP dissociation inhibitor (GDI)

[1] Schalk I, Zeng K, Wu SK, Stura EA, Matteson J, Huang M, Tandon A, Wilson IA, Balch WE, Nature 1996;381:42-48.

218. Oxidoreductase family (GFO_IDH_MocA)

This family of enzymes utilise NADP or NAD. This family: is called the GFO/IDH/MOCA family in swiss-prot.

[1] Kingston RL, Scopes RK, Baker EN, Structure 1996;4:1413-1428.

219. GHMP kinases putative ATP-binding domain

The following kinases contains, in their N-terminal section, a conserved Gly/Ser-rich region which is probably involved in the binding of ATP [1]. These kinases are listed below. - Galactokinase (EC 2.7.1.6). - Homoserine kinase (EC 2.7.1.39). - Mevalonate kinase (EC 2.7.1.36). - Phosphomevalonate kinase (EC 2.7.4.2). This group of kinases was called 'GHMP' (from the first letter of their substrate)

Consensus pattern: [LIVM]-[PK]-x-[GSTA]-x(0,1)-G-L-[GS]-S-S-[GSA]-[GSTAC]-

[1] Tsay Y.H., Robinson G.W. Mol. Cell. Biol. 11:620-631(1991).

220. Glucose inhibited division protein A family signatures (GIDA)

Bacterial glucose inhibited division protein A (gene gidA) is a protein of 70Kd whose function is not yet known and whose sequence is highly conserved. It is evolutionary related to yeast hypothetical protein YGL236C, Caenorhabditis elegans hypothetical protein F52H3.2 and a Bacillus subtilis protein called gid (and which is different from B.subtilis gidA). Two highly conserved regions were selected as signature patterns. Both regions are located in the central region of the protein.

Consensus pattern: [GS]-[PT]-x-Y-C-P-S-[LIVM]-E-x-K-[LIVM]-x-[KR]-

Consensus pattern: A-G-Q-x-[NT]-G-x(2)-G-Y-x-E-[SAG](3)-[QS]-G-[LIVM](2)-A-G-[LIVMT]-N-A-

221. (GLFV_dehydrog)

Glu / Leu / Phe / Val dehydrogenases active site

- Glutamate dehydrogenases (EC 1.4.1.2, EC 1.4.1.3, and EC 1.4.1.4) (GluDH) are enzymes that catalyze the NAD- or NADP-dependent reversible deamination of glutamate into alpha-ketoglutarate [1,2]. GluDH isozymes are generally involved with either ammonia assimilation or glutamate catabolism.
- Leucine dehydrogenase (EC 1.4.1.9) (LeuDH) is a NAD-dependent enzyme that catalyzes the reversible deamination of leucine and several other aliphatic amino acids to their keto analogues [3].
- Phenylalanine dehydrogenase (EC 1.4.1.20) (PheDH) is a NAD-dependent enzyme that catalyzes the reversible deamidation of L-phenylalanine into phenyl-pyruvate [4].
- Valine dehydrogenase (EC 1.4.1.8) (ValDH) is a NADP-dependent enzyme that catalyzes the reversible deamidation of L-valine into 3-methyl-2-oxobutanoate [5].

These dehydrogenases are structurally and functionally related. A conserved lysine residue located in a glycine-rich region has been implicated in the catalytic mechanism. The conservation of the region around this residue allows the derivation of a signature pattern for such type of enzymes.

Consensus pattern[LIV]-x(2)-G-G-[SAG]-K-x-[GV]-x(3)-[DNST]-[PL] [K is the active site residue] Sequences known to belong to this class detected by the pattern ALL.

Note all known sequences from this family have Pro in the last position of the pattern with the exception of yeast GluDH which as Leu.

[1] Britton K.L., Baker P.J., Rice D.W., Stillman T.J. Eur. J. Biochem. 209:851-859(1992).
[2] Benachenhou-Lahfa N., Forterre P., Labedan B. J. Mol. Evol. 36:335-346(1993).
[3] Nagata S., Tanizawa K., Esaki N., Sakamoto Y., Ohshima T., Tanaka H., Soda K. Biochemistry 27:9056-9062(1988).
[4] Takada H., Yoshimura T., Ohshima T., Esaki N., Soda K. J. Biochem. 109:371-376(1991).
[5] Hutchinson C.R., Tang L. J. Bacteriol. 175:4176-4185(1993).

222. GMC oxidoreductases signatures

The following FAD flavoproteins oxidoreductases have been found [1,2] to be evolutionary related. These enzymes, which are called 'GMC oxidoreductases', are listed below. - Glucose oxidase (EC 1.1.3.4) (GOX) from Aspergillus niger. Reaction catalyzed: glucose + oxygen -> delta-gluconolactone + hydrogen peroxide. - Methanol oxidase (EC 1.1.3.13) (MOX) from fungi. Reaction catalyzed: methanol + oxygen -> acetaldehyde + hydrogen peroxide. - Choline dehydrogenase (EC 1.1.99.1) (CHD) from bacteria. Reaction catalyzed: choline + unknown acceptor -> betaine acetaldehyde + reduced acceptor. - Glucose dehydrogenase (GLD) (EC 1.1.99.10) from Drosophila. Reaction catalyzed: glucose + unknown acceptor -> delta-gluconolactone + reduced acceptor. - Cholesterol oxidase (CHOD) (EC 1.1.3.6) from Brevibacterium sterolicum and Streptomyces strain SA-COO. Reaction catalyzed: cholesterol + oxygen -> cholest-4-en-3-one + hydrogen peroxide. - AlkJ [3], an alcohol dehydrogenase from Pseudomonas oleovorans, which converts aliphatic medium-chain-length alcohols into aldehydes. This family also includes a lyase: - (R)-mandelonitrile lyase (EC 4.1.2.10) (hydroxynitrile lyase) from plants [4], an enzyme involved in cyanogenis, the release of hydrogen cyanide from injured tissues. These enzymes are proteins of size ranging from 556 (CHD) to 664 (MOX) amino acid residues which share a number of regions of sequence similarities. One of these regions, located in the N-terminal section, corresponds to the FAD ADP-binding domain. The function of the other conserved domains is not yet known; two of these domains were selected as signature patterns. The first one is located in the N-terminal section of these enzymes, about 50 residues after the ADP-binding domain, while the second one is located in the central section.

Consensus pattern: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]- x(2)-[PAG]-x(5)-[DNESH]-

Consensus pattern: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G-

[1] Cavener D.R. J. Mol. Biol. 223:811-814(1992).
[2] Henikoff S., Henikoff J.G. Genomics 19:97-107(1994).
[3] van Beilen J.B., Eggink G., Enequist H., Bos R., Witholt B. Mol. Microbiol. 6:3121-3136(1992).
[4] Cheng I.P., Poulton J.E. Plant Cell Physiol. 34:1139-1143(1993).

223. (GMP_synt_C)

Glutamine amidotransferases class-I active site

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known as glutamine amidotransferase (GATase) (EC 2.4.2.-) [1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I (also known as trpG-type) and class-II (also known as purF-type). Class-I GATase domains have been found in the following enzymes:

- The second component of anthranilate synthase (AS) (EC 4.1.3.27) [4]. AS catalyzes the biosynthesis of anthranilate from chorismate and glutamine. AS is generally a dimeric enzyme: the first component can synthesize anthranilate using ammonia rather than glutamine, whereas component II provides the GATase activity. In some bacteria and in fungi the GATase component of AS is part of a multifunctional protein that also catalyzes other steps of the biosynthesis of tryptophan.
- The second component of 4-amino-4-deoxychorismate (ADC) synthase (EC 4.1.3. -), a dimeric prokaryotic enzyme that function in the pathway that catalyzes the biosynthesis of para-aminobenzoate (PABA) from chorismate and glutamine. The second component (gene pabA) provides the GATase activity [4].
- CTP synthase (EC 6.3.4.2). CTP synthase catalyzes the final reaction in the biosynthesis of pyrimidine, the ATP-dependent formation of CTP from UTP and glutamine. CTP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the C-terminal section [2].
- GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2). GMP synthase catalyzes the ATP-dependent formation of GMP from xanthosine 5'-phosphate and glutamine. GMP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the N-terminal section [5].
- Glutamine-dependent carbamoyl-phosphate synthase (EC 6.3.5.5) (GD-CPSase); an enzyme involved in both arginine and pyrimidine biosynthesis and which catalyzes the ATP-dependent formation of carbamoyl phosphate from glutamine and carbon dioxide. In bacteria

GD-CPSase is composed of two subunits: the large chain (gene carB) provides the CPSase activity, while the small chain (gene carA) provides the GATase activity. In yeast the enzyme involved in arginine biosynthesis is also composed of two subunits: CPA1 (GATase), and CPA2 (CPSase). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme (called URA2 in yeast, rudimentary in Drosophila, and CAD in mammals). The GATase domain is located at the N-terminal extremity of this polyprotein [6].

- Phosphoribosylformylglycinamidine synthase II (EC 6.3.5.3), an enzyme that catalyzes the fourth step in the de novo biosynthesis of purines. In some species of bacteria, FGAM synthase II is composed of two subunits: a small chain (gene purQ) which provides the GATase activity and a large chain (gene purL) which provides the aminator activity.
- The histidine amidotransferase hisH, an enzyme that catalyzes the fifth step in the biosynthesis of histidine in prokaryotes.

In the second component of AS a cysteine has been shown [7] to be essential for the amidotransferase activity. The sequence around this residue is well conserved in all the above GATase domains and can be used as a signature pattern for class-I GATase.

Consensus pattern[PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]- x-[LIVMFA] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for 6 sequences.

Note: in the first position of the pattern Pro is found in all cases except in the slime mold GD-CPSase where it is replaced by Ala.

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).
[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).
[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).
[4] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).
[5] Zalkin H., Argos P., Narayana S.V.L., Tiedeman A.A., Smith J.M. J. Biol. Chem. 260:3350-3354(1985).
[6] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).
[7] Tso J.Y., Hermodson M.A., Zalkin H. J. Biol. Chem. 255:1451-1457(1980).

224. Glutathione peroxidases signatures (GSHPx)

Glutathione peroxidase (EC 1.11.1.9) (GSHPx) [1,2] is an enzyme that catalyzes the reduction of hydroxyperoxides by glutathione. Its main function is to protect against the damaging effect of endogenously formed hydroxyperoxides. In higher vertebrates at least four forms of GSHPx are known to exist: a ubiquitous cytosolic form (GSHPx-1), a gastrointestinal cytosolic for (GSHPx-GI) [3], a plasma secreted form (GSHPx-P) [4], and a epididymal secretory form (GSHPx-EP). In addition to these characterized forms, the sequence of a protein of unknown function [5] has been shown to be evolutionary related to those of GSHPx's. In filarial nematode parasites such as Brugia pahangi the major soluble cuticular protein, known as gp29, is a secreted GSHPx which could provide a mechanism of resistance to the immune reaction of the mammalian host by neutralizing the products of the oxidative burst of leukocytes [6].Escherichia coli protein btuE, a periplasmic protein involved in the transport of vitamin B12, is also evolutionary related to GSHPx's; the significance of this relationship is not yet clear. Selenium, in the form of selenocysteine [7] is part of the catalytic site of GSHPx. The sequence around the selenocysteine residue is moderately well conserved in GSHPx's and the related proteins and can be used as a signature pattern. As a second signature for this family of proteins a highly conserved octapeptide located in the central section of these proteins was selected.

Consensus pattern: [GN]-[RKHNFYC]-x-[LIVMFC]-[LIVMF](2)-x-N-[VT]-x-[STC]-x-C-[GA]-x-T [C is the active site selenocysteine residue]

Consensus pattern: [LIV]-[AGD]-F-P-[CS]-[NG]-Q-

[1] Mannervik B. Meth. Enzymol. 113:490-495(1985).

[2] Mullenbach G.T., Tabrizi A., Irvine B.D., Bell G.I., Tainer J.A., Hallewell R.A. Protein Eng. 2:239-246(1988).

[3] Chu F.F., Doroshow J.H., Esworthy R.S. J. Biol. Chem. 268:2571-2576(1993).

[4] Takahashi K., Akasaka M., Yamamoto Y., Kobayashi C., Mizoguchi J., Koyama J. J. Biochem. 108:145-148(1990).

[5] Dunn D.K., Howells D.D., Richardson J., Goldfarb P.S. Nucleic Acids Res. 17:6390-6390(1989).

[6] Cookson E., Blaxter M.L., Selkirk M.E. Proc. Natl. Acad. Sci. U.S.A. 89:5837-5841(1992).

[7] Stadtman T.C. Annu. Rev. Biochem. 59:111-127(1990).

225. (GST)

Glutathione S-transferases

Function: conjugation of reduced glutathione to a variety of targets. Also included in the alignment, but are not GSTs S-crystallins from squid. Similarity to GST was previously noted. Eukaryotic elongation factors 1-gamma. Not known to have GST activity; similarity not previously recognized. Supported by HMM and manual alignment inspection. HSP26 family of stress-related proteins. including auxin-regulated proteins in plants and stringent starvation proteins in E. coli. Not known to have GST activity. Similarity not previously recognized. Supported by HMM and manual alignment inspection. Alignment spans entire protein.

226. GTP1/OBG family signature

A widespread family of GTP-binding proteins has been recently characterized [1,2]. This family currently includes: - Mouse and Xenopus protein DRG. - Human protein DRG2. - Drosophila protein 128up. - Fission yeast protein gtp1. - A Halobacterium cutirubrum hypothetical protein in a ribosomal protein gene cluster. - Bacillus subtilis protein obg. Obg has been experimentally shown to bind GTP. - Escherichia coli hypothetical protein yhbZ. - Haemophilus influenzae hypothetical protein HI0877. - Mycoplasma genitalium hypothetical protein MG384. - Yeast hypothetical protein YAL036c (FUN11). - Yeast hypothetical protein YGR173w. - Caenorhabditis elegans hypothetical protein C02F5.3.The function of the proteins that belong to this family is not yet known. They are polypeptides of about 40 to 48 Kd which contain the five small sequence elements characteristic of GTP-binding proteins [3]. As a signature pattern the region that correspond to the ATP/GTP B motif (also called G-3 inGTP-binding proteins) was selected.

Consensus pattern: D-[LIVM]-P-G-[LIVM](2)-[DEY]-[GN]-A-x(2)-G-x-G -

[1] Sazuka T., Tomooka Y., Ikawa Y., Noda M., Kumar S. Biochem. Biophys. Res. Commun. 189:363-370(1992).

[2] Hudson J.D., Young P.G. Gene 125:191-193(1993).

[3] Bourne H.R., Sanders D.A., McCormick F. Nature 349:117-127(1991).

227. (GTP_EFTU1)

ATP/GTP-binding site motif A (P-loop)

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].There are numerous ATP- or GTP-binding proteins in which the P-loop is found. Listed below are a number of protein families for which the relevance of the presence of such motif has been noted: - ATP synthase alpha and beta subunits (see <PDOC00137>). - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>). - Dynamins and dynamin-like proteins (see <PDOC00362>). - Guanylate kinase (see <PDOC00670>). - Thymidine kinase (see <PDOC00524>). - Thymidylate kinase (see <PDOC01034>). - Shikimate kinase (see <PDOC00868>). - Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>). - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>). - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran (see <PDOC00859>). - ADP-ribosylation factors family (see <PDOC00781>). - Bacterial dnaA protein (see <PDOC00771>). - Bacterial recA protein (see <PDOC00131>). - Bacterial recF protein (see <PDOC00539>). - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family (See <PDOC00388>). - Bacterial type II secretion system protein E (see <PDOC00567>).Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for adenylate

kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

-Consensus pattern: [AG]-x(4)-G-K-[ST]-

[1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).

[2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).

[3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).

[4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).

[5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).

[6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).

[7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).

[8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

GTP-binding elongation factors signature (GTP_EFTU2)

Elongation factors [1,2] are proteins catalyzing the elongation of peptide chains in protein biosynthesis. In both prokaryotes and eukaryotes, there are three distinct types of elongation factors, as described in the following table: -- Eukaryotes Prokaryotes Function -- EF-1alpha EF-Tu Binds GTP and an aminoacyl-tRNA; delivers the latter to the A site of ribosomes. EF-1beta EF-Ts Interacts with EF-1a/EF-Tu to displace GDP and thus allows the regeneration of GTP-EF-1a. EF-2 EF-G Binds GTP and peptidyl-tRNA and translocates the latter from the A site to the P site. ---The GTP-binding elongation factor family also includes the following proteins: - Eukaryotic peptide chain release factor GTP-binding subunits [3]. These proteins interact with release factors that bind to ribosomes that have encountered a stop codon at their decoding site and help them to induce release of the nascent polypeptide. The yeast protein was known as SUP2 (and also as SUP35, SUF12 or GST1) and the human homolog as

GST1-Hs. - Prokaryotic peptide chain release factor 3 (RF-3) (gene prfC). RF-3 is a class-II RF, a GTP-binding protein that interacts with class I RFs (see <PDOC00607>) and enhance their activity [4]. - Prokaryotic GTP-binding protein lepA and its homolog in yeast (gene GUF1) and in Caenorhabditis elegans (ZK1236.1). - Yeast HBS1 [5]. - Rat statin S1 [6], a protein of unknown function which is highly similar to EF-1alpha. - Prokaryotic selenocysteine-specific elongation factor selB [7], which seems to replace EF-Tu for the insertion of selenocysteine directed by the UGA codon. - The tetracycline resistance proteins tetM/tetO [8,9] from various bacteria such as Campylobacter jejuni, Enterococcus faecalis, Streptococcus mutans and Ureaplasma urealyticum. Tetracycline binds to the prokaryotic ribosomal 30S subunit and inhibits binding of aminoacyl-tRNAs. These proteins abolish the inhibitory effect of tetracycline on protein synthesis. - Rhizobium nodulation protein nodQ [10]. - Escherichia coli hypothetical protein yihK [11].In EF-1-alpha, a specific region has been shown [12] to be involved in a conformational change mediated by the hydrolysis of GTP to GDP. This region is conserved in both EF-1alpha/EF-Tu as well as EF-2/EF-G and thus seems typical for GTP-dependent proteins which bind non-initiator tRNAs to the ribosome. The pattern developed for this family of proteins include that conserved region.

Consensus pattern: D-[KRSTGANQFYW]-x(3)-E-[KRAQ]-x-[RKQD]-[GC]-[IVMK]-[ST]-[IV]-x(2)-[GSTACKRNQ]-

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

[2] Moldave K. Annu. Rev. Biochem. 54:1109-1149(1985).

[3] Stansfield I., Jones K.M., Kushnirov V.V., Dagkesamanskaya A.R., Poznyakovski A.I., Paushkin S.V., Nierras C.R., Cox B.S., Ter-Avanesyan M.D., Tuite M.F. EMBO J. 14:4365-4373(1995).

[4] Grentzmann G., Brechemier-Baey D., Heurgue-Hamard V., Buckingham R.H. J. Biol. Chem. 270:10595-10600(1995).

[5] Nelson R.J., Ziegelhoffer T., Nicolet C., Werner-Washburne M., Craig E.A. Cell 71:97-105(1992).

[6] Ann D.K., Moutsatsos I.K., Nakamura T., Lin H.H., Mao P.-L., Lee M.-J., Chin S., Liem R.K.H., Wang E. J. Biol. Chem. 266:10429-10437(1991).

[7] Forchammer K., Leinfeldr W., Bock A. Nature 342:453-456(1989).

[8] Manavathu E.K., Hiratsuka K., Taylor D.E. Gene 62:17-26(1988).

[9] Leblanc D.J., Lee L.N., Titmas B.M., Smith C.J., Tenover F.C. J. Bacteriol. 170:3618-3626(1988).

[10] Cervantes E., Sharma S.B., Maillet F., Vasse J., Truchet G., Rosenberg C. Mol. Microbiol. 3:745-755(1989).

[11] Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R. Nucleic Acids Res. 21:3391-3398(1993).

[12] Moller W., Schipper A., Amons R. Biochimie 69:983-989(1987).

228. GTP cyclohydrolase II.

GTP cyclohydrolase II catalyses the first committed step in the biosynthesis of riboflavin.

[1] Richter G, Ritz H, Katzenmeier G, Volk R, Kohnle A, Lottspeich F, Allendorf D, Bacher A, J Bacteriol 1993;175:4045-4051.

229. Galactose-1-phosphate uridyl transferase signatures (GalP_UDP_transf)

Galactose-1-phosphate uridyl transferase (EC 2.7.7.10) (galT) catalyzes the transfer of an uridyldiphosphate group on galactose (or glucose) 1-phosphate. During the reaction, the uridyl moiety links to a histidine residue. In the Escherichia coli enzyme, it has been shown [1] that two histidine residues separated by a single proline residue are essential for enzyme activity. On the basis of sequence similarities, two apparently unrelated families seem to exist. Class-I enzymes are found in eukaryotes as well as some bacteria such as Escherichia coli or Streptomyces lividans, while class-II enzymes have been found so far only in bacteria such as Bacillus subtilis or Lactobacillus helveticus [2]. Signature patterns for both families were developed. For class-I enzymes the signature is based on the active site residues. For class-II enzymes a region which also includes two conserved histidines was chosen.

Consensus pattern: F-E-N-[RK]-G-x(3)-G-x(4)-H-P-H-x-Q [The two H's are the active site residues]-

Consensus pattern: D-L-P-I-V-G-G-[ST]-[LIVM](2)-[SA]-H-[DEN]-H-[FY]-Q-G-G -

Note: class-I enzymes are structurally related to the HIT family of proteins (see

<PDOC00694

[1] Reichardt J.K.V., Berg P. Nucleic Acids Res. 16:9017-9026(1988).

[2] Mollet B., Pilloud N. J. Bacteriol. 173:4464-4473(1991).

230. Gamma-thionins family signature

The following small plant proteins are evolutionary related:

- Gamma-thionins from wheat endosperm (gamma-purothionins) and barley (gamma- hordothionins) which are toxic to animal cells and inhibit protein synthesis in cell free systems [1].
- A flower-specific thionin (FST) from tobacco [2].
- Antifungal proteins (AFP) from the seeds of Brassicaceae species such as radish, mustard, turnip and Arabidopsis thaliana [3].
- Inhibitors of insect alpha-amylases from sorghum [4].
- Probable protease inhibitor P322 from potato.
- A germination-related protein from cowpea [5].
- Anther-specific protein SF18 from sunflower [6]. SF18 is a protein that contains a gamma-thionin domain at its N-terminus and a proline-rich C- terminal domain.
- Soybean sulfur-rich protein SE60 [7].
- Vicia faba antibacterial peptides fabatin-1 and -2.

In their mature form, these proteins generally consist of about 45 to 50amino-acid residues. As shown in the following schematic representation, these peptides contain eight conserved cysteines involved in disulfide bonds.

```
+------------------------------------------+ | +-------------------+ | | | | |
xxCxxxxxxxxxxCxxxxxCxxxCxxxxxxxxxCxxxxxxCxCxxxC *****************|***| | |
+---|---------------+ | +------------------+
```

'C': conserved cysteine involved in a disulfide bond.

'*': position of the pattern.

Consensus pattern: [KRG]-x-C-x(3)-[SV]-x(2)-[FYWH]-x-[GF]-x-C-x(5)-C-x(3)-C [The four C's are involved in disulfide bonds]-

[1] Bruix M., Jimenez M.A., Santoro J., Gonzalez C., Colilla F.J., Mendez E., Rico M. Biochemistry 32:715-724(1993).

[2] Gu Q., Kawata E.E., Morse M.-J., Wu H.-M., Cheung A.Y. Mol. Gen. Genet. 234:89-96(1992).

[3] Terras F.R.G., Torrekens S., van Leuven F., Osborn R.W., Vanderleyden J., Cammue B.P.A., Broekaert W.F. FEBS Lett. 316:233-240(1993).

[4] Bloch C. Jr., Richardson M. FEBS Lett. 279:101-104(1991).

[5] Ishibashi N., Yamauchi D., Miniamikawa T. Plant Mol. Biol. 15:59-64(1990).

[7] Choi Y., Choi Y.D., Lee J.S. Plant Physiol. 101:699-700(1993).

231. Gelsolin. Gelsolin repeat. Number of members: 170

[1]Medline: 97433077. The crystal structure of plasma gelsolin: implications for actin severing, capping, and nucleation. Burtnick LD, Koepf EK, Grimes J, Jones EY, Stuart DI, McLaughlin PJ, Robinson RC; Cell 1997;90:661-670.

232. Germin family signature

Germins [1] are a family of homopentameric cereal glycoproteins expressed during germination which may play a role in altering the properties of cell walls during germinative growth. It has been shown that wheat and barleygermins act as oxalate oxidases (EC 1.2.3.4), an enzyme that catalyzes the oxidative degradation of oxalate to carbonate and hydrogen peroxide. Germins are highly similar to: - Germin-like proteins from various plants such as rape, violet or white mustard. - Slime mold spherulins 1a and 1b which are proteins that accumulate specifically during spherulation, a process induced by various forms of environmental stress which leads to encystment and dormancy. As a signature pattern the best conserved region was selected: a decapeptide located in the central section of these proteins.

Consensus pattern: G-x(4)-H-x-H-P-x-A-x-E-[LIVM]-

[1] Lane B.G. FASEB J. 8:294-301(1994).

233. (GlutR)

Glutamyl-tRNA reductase signature

Delta-aminolevulinic acid (ALA) is the obligatory precursor for the synthesis of all tetrapyrroles including porphyrin derivatives such as chlorophyll and heme. ALA can be synthesized via two different pathways: the Shemin (or C4) pathway which involves the single step condensation of succinyl-CoA and glycine and which is catalyzed by ALA synthase (EC 2.3.1.37) and via the C5pathway from the five-carbon skeleton of glutamate. The C5 pathway operates in the chloroplast of plants and algae, in cyanobacteria, in some eubacteria and in archaebacteria.

The initial step in the C5 pathway is carried out by glutamyl-tRNA reductase (GluTR) [1] which catalyzes the NADP-dependent conversion of glutamate- tRNA(Glu) to glutamate-1-semialdehyde (GSA) with the concomitant release of tRNA(Glu) which can then be recharged with glutamate by glutamyl-tRNA synthetase.

GluTR is a protein of about 50 Kd (467 to 550 residues) which contains a few conserved region. The best conserved region is located in positions 99 to 122 in the sequence of known GluTR. This region seems important for the activity of the enzyme. We have developed a signature pattern from that conserved region.

Consensus patternH-[LIVM]-x(2)-[LIVM]-[GSTAC](3)-[LIVM]-[DEQ]-S-[LIVMA]-[LIVM](2)-[GF]-E-x-[EQR]-[IV]-[LIT]-[STAG]-Q-[LIVM]-[KR] Sequences known to belong to this class detected by the pattern ALL.

[1] Jahn D., Verkamp E., Soell D. Trends Biochem. Sci. 17:215-218(1992).

234. (Glycoprotease)

Glycoprotease family signature (aka Peptidase_M22)

Glycoprotease (GCP) (EC 3.4.24.57) [1], or o-syaloglycoprotein endopeptidase, is a metalloprotease secreted by Pasteurella haemolytica which specifically cleaves O-sialoglycoproteins such as glycophorin A. The sequence of GCP is highly similar to the following uncharacterized proteins:

- Escherichia coli hypothetical protein ygjD (ORF-X).
- Bacillus subtilis hypothetical protein ydiE.
- Mycobacterium leprae hypothetical protein U229E.
- Mycobacterium tuberculosis hypothetical protein MtCY78.10.
- Synechocystis strain PCC 6803 hypothetical protein slr0807.
- Methanococcus jannaschii hypothetical protein MJ1130.
- Haloarcula marismortui hypothetical protein in HSH 3' region.
- Yeast hypothetical protein YKR038c.
- Yeast hypothetical protein QRI7.

One of the conserved regions contains two conserved histidines. It is possible that this region is involved in coordinating a metal ion such as zinc.

Consensus pattern[KR]-[GSAT]-x(4)-[FYWLH]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H- x(2)-[AG]-H-[LIVM] Sequences known to belong to this class detected by the pattern ALL.

Note: these proteins belong to family M22 in the classification of peptidases [2,E1].

[1] Abdullah K.M., Lo R.Y.C., Mellors A. J. Bacteriol. 173:5597-5603(1991).

[2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

235. (Glucosamine_iso)

Glucosamine/galactosamine-6-phosphate isomerases signature

Glucosamine-6-phosphate isomerase (EC 5.3.1.10) (or Glc-6-P deaminase) is the enzyme responsible for the conversion of glucosamine 6-phosphate into fructose6 phosphate [1]. It is the last specific step in the pathway for N-acetylglucosamine (GlcNAC) utilization in bacteria such as Escherichia coli (gene nagB) or in fungi such as Candida albicans (gene NAG1).Glc-6-P isomerase is evolutionary related to: - A putative Escherichia coli galactosamine-6-phosphate isomerase (gene agaI) [2]. - Escherichia coli hypothetical protein yieK. - Bacillus subtilis hypothetical protein ybfT. As a signature pattern a conserved region located in the central part of these enzymes was selected. This region contains a conserved histidine which

has been shown [1], in nagB, to be important for the pyranose ring-opening step of the catalytic mechanism

Consensus pattern: [LIVM]-x(3)-G-x-[LIT]-x-[LIV]-x-[LIVM]-x-G-[LIVM]-G-x- [DEN]-G-H-

[1] Oliva G., Fontes M.R.M., Garratt R.C., Altamirano M.M., Calcagno M.L., Horjales E. Structure 3:1323-1332(1995).

[2] Reizer J., Ramseier T.M., Reizer A., Charbit A., Saier M.H. Jr. Microbiology 142:231-250(1996).

236. Pneumovirus attachment glycoprotein G (glycoprotein G)

This family includes attachment proteins from respiratory synctial virus. Glycoprotein G has not been shown to have any neuraminidase or hemagglutinin activity (Swiss-Prot). The amino terminus is thought to be cytoplasmic, and the carboxyl terminus extracellular. The extracellular region contains four completely conserved cysteine residues.

[1] Johnson PR, Spriggs MK, Olmsted RA, Collins PL, Proc Natl Acad Sci U S A 1987;84:5625-5629.

237. Glycosyl transferases group 1

Mutations in this domain of Swiss:P37287 lead to disease (Paroxysmal Nocturnal haemoglobinuria). Members of this family transfer activated sugars to a variety of substrates, including glycogen, Fructose-6-phosphate and lipopolysaccharides. Members of this family transfer UDP, ADP, GDP or CMP linked sugars. The eukaryotic glycogen synthases may be distant members of this family.

238. Glycosyl transferases (Glycos_transf_2)

Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl-galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids.

239. (Glucos_transf_3)
Thymidine and pyrimidine-nucleoside phosphorylases signature

Thymidine phosphorylase (EC 2.4.2.4) catalyzes the reversible phosphorolysis of thymidine, deoxyuridine and their analogues to their respective bases and 2-deoxyribose 1-phosphate. This enzyme regulates the availability of thymidine and is therefore essential to nucleic acid metabolism.

In Escherichia coli (gene deoA), the enzyme is a dimer of identical subunits of about 48 Kd [1]. In humans it was first identified as platelet-derived endothelial cell growth factor (PD-ECGF) [E1] before being recognized [2] as thymidine phosphorylase.

Bacterial pyrimidine-nucleoside phosphorylase (EC 2.4.2.2) (gene pdp) [3] is an enzyme evolutionary and structurally related to thymidine phosphorylase.

A a well conserved region of 19 residues located in the N-terminal part of these proteins signature pattern for these enzymes was selected.

Consensus patternS-[GS]-R-[GA]-[LIV]-x(2)-[TA]-[GA]-G-T-x-D-x-[LIV]-E Sequences known to belong to this class detected by the pattern ALL.

[1] Walter M.R., Cook W.J., Cole L.B., Short S.A., Koszalka G.W., Krenitsky T.A., Ealick S.E. J. Biol. Chem. 265:14016-14022(1990).
[2] Furukawa T., Yoshimura A., Sumizawa T., Haraguchi M., Akiyama S.-I., Fukui K., Yamada Y. Nature 356:668-668(1992).
[3] Saxild H.H., Andersen L.N., Hammer K. J. Bacteriol. 178:424-434(1996).

240. Glycos_transf_4. Glycosyl transferase. Number of members: 44.

[1] Medline: 95252686. A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases. Lehrman MA; Glycobiology 1994;4:768-771.

241. Glycosyl hydrolases family 15. 21 members.

242. Glycosyl hydrolases family 16 signature

It has been shown [1] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities: - Bacterial beta-1,3-1,4-glucanases, or lichenases, (EC 3.2.1.73) mainly from Bacillus but also from Clostridium thermocellum (gene licB), Fibrobacter succinogenes and Rhodothermus marinus (gene bglA). - Bacillus circulans beta-1,3-glucanase A1 (EC 3.2.1.39) (gene glcA). - Lamarinase (EC 3.2.1.6) from Clostridium thermocellum (gene lam1). - Streptomyces coelicolor agarase (EC 3.2.1.81) (gene dagA). - Alteromonas carrageenovora kappa-carrageenase (EC 3.2.1.83) (gene cgkA).Two closely clustered conserved glutamates have been shown [2] to be involved in the catalytic activity of Bacillus licheniformis lichenase. The region was used that contains these residues as a signature pattern.

Consensus pattern: E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA] [The two E's are active site residues]-

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Juncosa M., Pons J., Dot T., Querol E., Planas A. J. Biol. Chem. 269:14530-14535(1994).

243. Glycosyl hydrolases family 17 signature

It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities: - Glucan endo-1,3-beta-glucosidases (EC 3.2.1.39) (endo-(1->3)-beta- glucanase) from various plants. This enzyme may be involved in the defense of plants against pathogens through its ability to degrade fungal cell wall polysaccharides. - Glucan 1,3-beta-glucosidase (EC 3.2.1.58) (exo-(1->3)-beta-glucanase) from yeast (gene BGL2). This enzyme may play a role in cell expansion during growth, in cell-cell fusion during mating, and in spore release during sporulation. - Lichenases (EC 3.2.1.73) (endo-(1->3,1->4)-beta-glucanase) from various plants. The best conserved region in the sequence of these enzymes is located in their central section. This region contains a

conserved tryptophan residue which could be involved in the interaction with the glucan substrates [2] and it also contains a conserved glutamate which has been shown [3] to act as the nucleophile in the catalytic mechanism. this region was used as a signature pattern.

Consensus pattern: [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ] [E is an active site residue]-

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Ori N., Sessa G., Lotan T., Himmelhoch S., Fluhr R. EMBO J. 9:3429-3436(1990).
[3] Varghese J.N., Garrett T.P.J., Colman P.M., Chen L., Hoj P.J., Fincher G.B. Proc. Natl. Acad. Sci. U.S.A. 91:2785-2789(1994).

244. Glyoxalase I signatures

Glyoxalase I (EC 4.4.1.5) (lactoylglutathione lyase) catalyzes the first step of the glyoxal pathway, the transformation of methylglyoxal and glutathioneinto S-lactoylglutathione which is then converted by glyoxalase II to lactic acid [1]. Glyoxalase I is an ubiquitous enzyme which binds one mole of zinc per subunit. The bacterial and yeast enzymes are monomeric while the mammalian one is homodimeric. The sequence of glyoxalase I is well conserved. In bacteria and mammals, the enzyme is a protein of about 130 to 180 residues while in fungi it is about twice longer. In these organisms the enzyme is built out of the tandem repeat of an homologous domain. Two signature patterns for this family were derived. The first one is located in the N-terminal region while the second one is located in the central section of the protein and contains a conserved histidine that could be implicated in the binding of the zinc atom.

Consensus pattern: [HQ]-[IVT]-x-[LIVFY]-x-[IV]-x(5)-[STA]-x(2)-F-[YM]-x(2,3)- [LMF]-G-[LMF]-
Consensus pattern: G-[NTKQ]-x(0,5)-[GA]-[LVFY]-[GH]-H-[IVF]-[CGA]-x-[STAGLE]-x(2)-[DNC]-

[1] Kim N.-S., Umezawa Y., Ohmura S., Kato S. J. Biol. Chem. 268:11217-11221(1993).

245. (Glypican)

Glypicans signature

Glypicans [1,2] are a family of heparan sulfate proteoglycans which are anchored to cell membranes by a glycosylphosphatidylinositol (GPI) linkage. Structurally, these proteins consist of three separate domains:

a) A signal sequence;
b) An extracellular domain of about 500 residues that contains 12 conserved cysteines probably involved in disulfide bonds and which also contains the sites of attachment of the heparan sulfate glycosaminoglycan side chains;
c) A C-terminal hydrophobic region which is post-translationally removed after formation of the GPI-anchor.

The proteins known to belong to this family are:

- Glypican 1 (GPC1).
- Glypican 2 (GPC2) or cerebroglycan.
- Glypican 3 (GPC3) or OCI-5. In man, defects in GPC3 are the cause of a X-linked genetic disease, Simpson-Galabi-Behmel syndrome (SGBS).
- K-glypican.
- Glypican 5 (GPC5).
- Drosophila protein dally.

The signature pattern that was developed for glypicans is located in the central section of the extracellular domain and contains five of the conserved cysteines.

Consensus patternC-x(2)-C-x-G-[LIVM]-x(4)-P-C-x(2)-[FY]-C-x(2)-[LIVM]-x(2)- G-C [The C's are probably involved in a disulfide bonds] Sequences known to belong to this class detected by the pattern ALL, except for dally.

[1] Weksberg R., Squire J.A., Templeton D.M. Nat. Genet. 12:225-227(1996).
[2] Watanabe K., Yamada H., Yamaguchi Y. J. Cell Biol. 130:1207-1218(1995).

246. Granins signatures

Granins (chromogranins or secretogranins) [1] are a family of acidic proteins present in the secretory granules of a wide variety of endocrine and neuro-endocrine cells. The exact function(s) of these proteins is not yet known but they seem to be the precursors of biologically active peptides and/or they may act as helper proteins in the packaging of peptide hormones and neuropeptides. Three members of this family of proteins show some sequence similarities: - Chromogranin A (CGA) [2]. CGA is a protein of about 420 residues; it is the precursor of the peptide pancreastatin which strongly inhibits glucose- induced insulin release from the pancreas. - Secretogranin 1 (chromogranin B). A sulfated protein of about 600 residues. - Secretogranin 2 (chromogranin C). A sulfated protein of about 650 residues. Apart from their subcellular location and the abundance of acidic residues(Asp and Glu), these proteins do not share many structural similarities. Only one short region, located in the C-terminal section, is conserved in all these proteins. Chromogranins A and B share a region of high similarity in their N-terminal section; this region includes two cysteine residues involved in a disulfide bond

Consensus pattern: [DE]-[SN]-L-[SAN]-x(2)-[DE]-x-E-L-

Consensus pattern: C-[LIVM](2)-E-[LIVM](2)-S-[DN]-[STA]-L-x-K-x-S-x(3)- [LIVM]-[STA]-x-E-C [The two C's are linked by a disulfide bond]-

[1] Huttner W.B., Gerdes H.-H., Rosa P. Trends Biochem. Sci. 16:27-30(1991).

[2] Simon J.-P., Aunis D. Biochem. J. 262:1-13(1989).

247. grpE protein signature

In prokaryotes the grpE protein [1] stimulates, jointly with dnaJ, the ATPase activity of the dnaK chaperone. It seems to accelerate the release of ADP from dnaK thus allowing dnaK to recycle more efficiently. GrpE is a protein of about 22 to 25 Kd. In yeast, an evolutionary related mitochondrial protein(gene GRPE) has been shown [2] to associate with the mitochondrial hsp70protein and to thus play a role in the import of proteins from the cytoplasm. As a signature pattern, the most conserved region of grpE was selected. It is located in the C-terminal section.

Consensus pattern: [FL]-[DN]-[PHEA]-x(2)-[HM]-x-A-[LIVMTN]-x(16,20)-G-[FY]- x(3)-[DEG]-x(2)-[LIVM]-[RI]-x-[SA]-x-V-x-[IV]-

[1] Georgopoulos C., Welch W. Annu. Rev. Cell Biol. 9:601-635(1993).

[2] Bolliger L., Deloche O., Glick B.S., Georgopoulos C., Jenoe P., Kronidou N., Horst M., Morishima N., Schatz G. EMBO J. 13:1998-2006(1994).

248. Guanylate kinase signature and profile

Guanylate kinase (EC 2.7.4.8) (GK) [1] catalyzes the ATP-dependent phosphorylation of GMP into GDP. It is essential for recycling GMP and indirectly, cGMP. In prokaryotes (such as Escherichia coli), lower eukaryotes (such as yeast) and in vertebrates, GK is a highly conserved monomeric protein of about 200 amino acids. GK has been shown [2,3,4] to be structurally similar to the following proteins: - Protein A57R (or SalG2R) from various strains of Vaccinia virus. This protein is highly similar to GK, but contains a frameshift mutation in the N-terminal section and could therefore be inactive in that virus. The following proteins are characterized by the presence in their sequence of one or more copies of the DHR domain, a SH3 domain (see <PDOC50002> as well as a C-terminal GK-like domain, these protein are collectively termed MAGUKs (membrane-associated guanylate kinase homologs) [5]: - Drosophila lethal(1)discs large-1 tumor suppressor protein (gene dlg1). This protein is associated with septate junctions in developing flies and defects in the dlg1 gene cause neoplastic overgrowth of the imaginal disks. - Mammalian tight junction protein Zo-1. - A family of mammalian synaptic proteins that seem to interact with the cytoplasmic tail of NMDA receptor subunits. This family currently consist of SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1 and SAP102. - Vertebrate 55 Kd erythrocyte membrane protein (p55). p55 is a palmitoylated, membrane-associated protein of unknown function. - Caenorhabditis elegans protein lin-2, which may play a structural role in the induction of the vulva. - Rat protein CASK. - Human protein DLG2. - Human protein DLG3.There is an ATP-binding site (P-loop) in the N-terminal section of GK. This region is not conserved in the GK-like domain of the above proteins which are therefore unlikely to be kinases. However these proteins retain the residues known, in GK, to be involved in the binding of GMP. As a signature pattern a highly conserved region was selected that contains two arginine and a tyrosine which are involved in GMP-binding

Consensus pattern: T-[ST]-R-x(2)-[KR]-x(2)-[DE]-x(2)-G-x(2)-Y-x-[FY]-[LIVMK]-

[1] Stehle T., Schulz G.E. J. Mol. Biol. 224:1127-1141(1992).
[2] Bryant P.J., Woods D.F. Cell 68:621-622(1992).
[3] Goebl M.G. Trends Biochem. Sci. 17:99-99(1992).
[4] Zschocke P.D., Schiltz E., Schulz G.E. Eur. J. Biochem. 213:263-269(1993).
[5] Woods D.F., Bryant P.J. Mech. Dev. 44:85-89(1994).

249. (Glyco_hydro_35)
Glycosyl hydrolases family 35 putative active site

Beta-galactosidases (EC 3.2.1.23) from mammals, fungi, plants and the bacteria Xanthomonas manihotis are evolutionary related [1,2]. They belong to family 35 in the classification of glycosyl hydrolases [3,E1].

Mammalian beta-galactosidase is a lysosomal enzyme (gene GLB1) which cleaves the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans and whose deficiency is the cause of the genetic disease Gm(1) gangliosidosis (Morquio disease type B).

On of the best conserved regions in these enzymes contains a glutamic acid residue which, on the basis of similarities with other families of glycosyl hydrolases [4], probably acts as the proton donor in the catalytic mechanism. This region wss used as a signature pattern.

Consensus pattern: G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY] [The second E is the putative active site residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P. Glycobiology 5:603-610(1995).
[2] Carey A.T., Holt K., Picard S., Wilde R., Tucker G.A., Bird C.R., Schuch W., Seymour G.B. Plant Physiol. 108:1099-1107(1995).
[3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993).
[4] Henrissat B., Callebaut I., Fabrega S., Lehn P., Mornon J.-P., Davies G. Proc. Natl. Acad. Sci. U.S.A. 92:7090-7094(1995).

250. (Glyco_hydro_16)
Glycosyl hydrolases family 16 signature

It has been shown [1] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- Bacterial beta-1,3-1,4-glucanases, or lichenases, (EC 3.2.1.73) mainly from Bacillus but also from Clostridium thermocellum (gene licB), Fibrobacter succinogenes and Rhodothermus marinus (gene bglA).
- Bacillus circulans beta-1,3-glucanase A1 (EC 3.2.1.39) (gene glcA).
- Lamarinase (EC 3.2.1.6) from Clostridium thermocellum (gene lam1).
- Streptomyces coelicolor agarase (EC 3.2.1.81) (gene dagA).
- Alteromonas carrageenovora kappa-carrageenase (EC 3.2.1.83) (gene cgkA).

Two closely clustered conserved glutamates have been shown [2] to be involved in the catalytic activity of Bacillus licheniformis lichenase. The region that contains these residues as a signature pattern was used.

Consensus pattern E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA] [The two E's are active site residues]

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Juncosa M., Pons J., Dot T., Querol E., Planas A. J. Biol. Chem. 269:14530-14535(1994).

251. (Glyco_hydro_17)
Glycosyl hydrolases family 17 signature
(aka glycosyl_hydro4)

It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- Glucan endo-1,3-beta-glucosidases (EC 3.2.1.39) (endo-(1->3)-beta-glucanase) from various plants. This enzyme may be involved in the defense of plants against pathogens through its ability to degrade fungal cell wall polysaccharides.
- Glucan 1,3-beta-glucosidase (EC 3.2.1.58) (exo-(1->3)-beta-glucanase) from yeast (gene BGL2). This enzyme may play a role in cell expansion during growth, in cell-cell fusion during mating, and in spore release during sporulation.
- Lichenases (EC 3.2.1.73) (endo-(1->3,1->4)-beta-glucanase) from various plants.

The best conserved region in the sequence of these enzymes is located in their central section. This region contains a conserved tryptophan residue which could be involved in the interaction with the glucan substrates [2] and it also contains a conserved glutamate which has been shown [3] to act as the nucleophile in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ] [E is an active site residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Ori N., Sessa G., Lotan T., Himmelhoch S., Fluhr R. EMBO J. 9:3429-3436(1990).
[3] Varghese J.N., Garrett T.P.J., Colman P.M., Chen L., Hoj P.J., Fincher G.B. Proc. Natl. Acad. Sci. U.S.A. 91:2785-2789(1994).

252. (Glyco_hydro_3)
Glycosyl hydrolases family 3 active site

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Beta glucosidases (EC 3.2.1.21) from the fungi Aspergillus wentii (A-3), Hansenula anomala, Kluyveromyces fragilis, Saccharomycopsis fibuligera, (BGL1 and BGL2), Schizophyllum commune and Trichoderma reesei (BGL1).
- Beta glucosidases from the bacteria Agrobacterium tumefaciens (Cbg1),

Butyrivibrio fibrisolvens (bglA), Clostridium thermocellum (bglB), Escherichia coli (bglX), Erwinia chrysanthemi (bgxA) and Ruminococcus albus.

- Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52).
- Bacillus subtilis hypothetical protein yzbA.
- Escherichica coli hypothetical protein ycfO and HI0959, the corresponding Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus wentii beta- glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern[LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]- [ST]-D-x(2)-[SGADNI] [D is the active site residue] Sequences known to belong to this class detected by the patternALL.

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).
[3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).

253. (Glyco_hydro_28)
Polygalacturonase active site (aka PG)

Polygalacturonase (EC 3.2.1.15) (PG) (pectinase) [1,2] catalyzes the random hydrolysis of 1,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans. In fruit, polygalacturonase plays an important role in cell wall metabolism during ripening. In plant bacterial pathogens such as Erwinia carotovora or Pseudomonas solanacearum and fungal pathogens such as Aspergillus niger, polygalacturonase is involved in maceration and soft-rotting of plant tissue.

Exo-poly-alpha-D-galacturonosidase (EC 3.2.1.82) (exoPG) [3] hydrolyzes peptic acid from the non-reducing end, releasing digalacturonate.

Prokaryotic, eukaryotic PG and exoPG share a few regions of sequence similarity. The best conserved of these regions was selected. It is centered on a conserved histidine most probably involved in the catalytic mechanism [4].

Consensus pattern[GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)- [LIVM]-G-S [H is the putative active site residue] Sequences known to belong to this class detected by the patternALL.

Note: these proteins belong to family 28 in the classification of glycosyl hydrolases [5].

[1] Ruttowski E., Labitzke R., Khanh N.Q., Loeffler F., Gottschalk M., Jany K.-D. Biochim. Biophys. Acta 1087:104-106(1990).
[2] Huang J., Schell M.A. J. Bacteriol. 172:3879-3887(1990).
[3] He S.Y., Collmer A. J. Bacteriol. 172:4988-4995(1990).
[4] Bussink H.J.D., Buxton F.P., Visser J. Curr. Genet. 19:467-474(1991).
[5] Henrissat B. Biochem. J. 280:309-316(1991).

254. (Glyco_hydro_32)
Glycosyl hydrolases family 32 active site

It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- Inulinase (EC 3.2.1.7) (or inulase) from the fungi Kluyveromyces marxianus.
- Beta-fructofuranosidase (EC 3.2.1.26), commonly known as invertase in fungi and plants and as sucrase in bacteria (gene sacA or scrB).
- Raffinose invertase (EC 3.2.1.26) (gene rafD) from Escherichia coli plasmid pRSD2.
- Levanase (EC 3.2.1.65) (gene sacC) from Bacillus subtilis.

One of the conserved regions in these enzymes is located in the N-terminal section and contains an aspartic acid residue which has been shown [3], in yeast invertase to be important for the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G [D is the active site residue]
Sequences known to belong to this class detected by the patternALL.

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Gunasekaran P., Karunakaran T., Cami B., Mukundan A.G., Preziosi L., Baratti J. J. Bacteriol. 172:6727-6735(1990).
[3] Reddy V.A., Maley F. J. Biol. Chem. 265:10817-10120(1990).

255. (Glyco_hydro_1)
Glycosyl hydrolases family 1 signatures

It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Beta-glucosidases (EC 3.2.1.21) from various bacteria such as Agrobacterium strain ATCC 21400, Bacillus polymyxa, and Caldocellum saccharolyticum.
- Two plants (clover) beta-glucosidases (EC 3.2.1.21).
- Two different beta-galactosidases (EC 3.2.1.23) from the archaebacteria Sulfolobus solfataricus (genes bgaS and lacS).
- 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as Lactobacillus casei, Lactococcus lactis, and Staphylococcus aureus.
- 6-phospho-beta-glucosidases (EC 3.2.1.86) from Escherichia coli (genes bglB and ascB) and from Erwinia chrysanthemi (gene arbB).
- Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase).
- Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62). LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is evolutionary related to the above glycosyl hydrolases.

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-glucosidase from Agrobacterium, to be directly

involved in glycosidic bond cleavage by acting as a nucleophile. This region was used as a signature pattern. As a second signature pattern we selected a conserved region, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue.

Consensus pattern[LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue] Sequences known to belong to this class detected by the patternALL.

Note: this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].

Consensus patternF-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA] Sequences known to belong to this class detected by the pattern ALL.

Note: this pattern will pick up the last three domains of LPH.

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991).
[3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990).
[4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992).
[5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).

256. Glyco_hydro_20
Glycosyl hydrolase family 20
Previous Pfam IDs: glycosyl_hydr11;
Number of members: 33

257. (Glyco_hydro_9)
Glycosyl hydrolases family 9 active sites signatures
(aka Glycosyl_hydr12)

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family E [3] or as the glycosyl hydrolases family 9 [4,E1]. The enzymes which are currently known to belong to this family are listed below.

- Butyrivibrio fibrisolvens cellodextrinase 1 (ced1).
- Cellulomonas fimi endoglucanases B (cenB) and C (cenC).
- Clostridium cellulolyticum endoglucanase G (celCCG).
- Clostridium cellulovorans endoglucanase C (engC).
- Clostridium stercoararium endoglucanase Z (avicelase I) (celZ).
- Clostridium thermocellum endoglucanases D (celD), F (celF) and I (celI).
- Fibrobacter succinogenes endoglucanase A (endA).
- Pseudomonas fluorescens endoglucanase A (celA).
- Streptomyces reticuli endoglucanase 1 (cel1).
- Thermomonospora fusca endoglucanase E-4 (celD).

- Dictyostelium discoideum spore germination specific endoglucanase 270-6. This slime mold enzyme may digest the spore cell wall during germination, to release the enclosed amoeba.
- Endoglucanases from plants such as Avocado or French bean. In plants this enzyme may be involved the fruit ripening process.

Two of the most conserved regions in these enzymes are centered on conserved residues which have been shown [5,6], in the endoglucanase D from Cellulomonas thermocellum, to be important for the catalytic activity. The first region contains an active site histidine and the second region contains two catalytically important residues: an aspartate and a glutamate. Both regions were used as signature patterns.

Consensus pattern [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Cellulomonas fimi cenC and Streptomyces reticuli cel1.

Consensus pattern [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA] [D and E are active site residues] Sequences known to belong to this class detected by the pattern ALL, except for Fibrobacter succinogenes endA whose sequence seems to be incorrect.

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).
[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).
[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).
[4] Henrissat B. Biochem. J. 280:309-316(1991).
[5] Tomme P., Chauvaux S., Beguin P., Millet J., Aubert J.-P., Claeyssens M. J. Biol. Chem. 266:10313-10318(1991).
[6] Tomme P., van Beeumen J., Claeyssens M. Biochem. J. 285:319-324(1992).

258. Matrix protein (MA), p15 (GAG_ma)

The matrix protein, p15, is encoded by the gag gene. MA is involved in pathogenicity [1].

[1] : Pozsgay JM, Beilharz MW, Wines BD, Hess AD, Pitha PM, J Virol 1993;67:5989-5999.

259. Gag polyprotein, inner coat protein p12 (GAG_P12)

The retroviral p12 is a virion structural protein. p12 is proline rich. The function carried out by p12 in assembly and replication is unknown. p12C is associated with pathogenicity of the virus

[1] Pozsgay JM, Beilharz MW, Wines BD, Hess AD, Pitha PM, J Virol 1993;67:5989-5999.

260. Glutamine synthetase signatures (GLN-SYNT)

Glutamine synthetase (EC 6.3.1.2) (GS) [1] plays an essential role in the metabolism of nitrogen by catalyzing the condensation of glutamate and ammonia to form glutamine. There seem to be three different classes of GS [2,3,4]: - Class I enzymes (GSI) are specific to prokaryotes, and are oligomers of 12 identical subunits. The activity of GSI-type enzyme is controlled by the adenylation of a tyrosine residue. The adenylated enzyme is inactive. - Class II enzymes (GSII) are found in eukaryotes and in bacteria belonging to the Rhizobiaceae, Frankiaceae, and Streptomycetaceae families (these bacteria have also a class-I GS). GSII are octamer of identical subunits. Plants have two or more isozymes of GSII, one of the isozymes is translocated into the chloroplast. - Class III enzymes (GSIII) has, currently, only been found in Bacteroides fragilis and in butyrivibrio fibrisolvens. It is a hexamer of identical chains. It is much larger (about 700 amino acids) than the GSI (450 to 470 amino acids) or GSII (350 to 420 amino acids) enzymes. While the three classes of GS's are clearly structurally related, the sequence similarities are not so extensive. As signature patterns three conserved regions were selected. The first pattern is based on a conserved tetrapeptide in the N-terminal section of the enzyme, the second one is based on a glycine-rich region which is thought to be involved in ATP-binding. The third pattern is specific to class I glutamine synthetases and includes the tyrosine residue which is reversibly adenylated.

Consensus pattern: [FYWL]-D-G-S-S-x(6,8)-[DENQSTAK]-[SA]-[DE]-x(2)-[LIVMFY]-
Consensus pattern: K-P-[LIVMFYA]-x(3,5)-[NPAT]-G-[GSTAN]-G-x-H-x(3)-S-
Consensus pattern: K-[LIVM]-x(5)-[LIVMA]-D-[RK]-[DN]-[LI]-Y [Y is the site of adenylation]-

[1] Eisenberg D., Almassy R.J., Janson C.A., Chapman M.S., Suh S.W., Cascio D., Smith W.W. Cold Spring Harbor Symp. Quant. Biol. 52:483-490(1987).
[2] Kumada Y., Benson D.R., Hillemann D., Hosted T.J., Rochefort D.A., Thompson C.J., Wohlleben W., Tateno Y. Proc. Natl. Acad. Sci. U.S.A. 90:3009-3013(1993).
[3] Shatters R.G., Kahn M.L. J. Mol. Evol. 29:422-428(1989).
[4] Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F. J. Mol. Evol. 38:566-576(1994).

261. Globins profile (globin1)

Globins are heme-containing proteins involved in binding and/or transporting oxygen [1]. They belong to a very large and well studied family which is widely distributed in many organisms. The major groups of globins are: - Hemoglobins (Hb) from vertebrates. Hb is the protein responsible for transporting oxygen from the lungs to other tissues. It is a tetramer of two alpha and two beta chains. Most vertebrate species also express specific embryonic or fetal forms of hemoglobin where the alpha or the beta chains are replaced by a chain with higher oxygen affinity, as for the gamma, delta, epsilon and zeta chains in mammals, for example. - Myoglobins (Mg) from vertebrates. Mg is a monomeric protein responsible for oxygen storage in muscles. - Invertebrate globins [2]. A wide variety of globins are found in invertebrates. Molluscs generally have one or two muscle globins which are either monomeric or dimeric. Insects, such as the midge Chironomus thummi, have a large set of extracellular globins. Nematodes and annelids have a variety of intracellular and extracellular globins; some of them are multi- domain polypeptides (from two up to nine-domain globins) and some produce large, disulfide-bonded aggregates. - Leghemoglobins (Lg) from the root nodules of leguminous plants. Lg provides oxygen for bacteroids. - Flavohemoproteins from bacteria (Escherichia coli hmpA) and fungi [3]. These proteins consist of two distinct domains: an N-terminal globin domain and a C-terminal FAD-containing reductase domain. In bacteria such as Vitreoscilla, the enzyme-associated globin is a single domain protein. All these globins seem to have evolved from a common ancestor. The profile developed to detect members of the globin family is based on a structural alignment of selected globin sequences [1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).[2] Goodman M., Pedwaydon J., Czelusniak J., Suzuki T., Gotoh T., Moens L., Shishikura F., Walz D., Vinogradov S. J. Mol. Evol. 27:236-249(1988).

Plant hemoglobins signature (globin2)

Leghemoglobins [1] are hemoproteins present in the root nodules of leguminousplants. Leghemoglobins are structurally and functionally related to hemoglobin and myoglobin. By providing oxygen to the bacteroids, they are essential for symbiotic nitrogen fixation. Structurally related hemoglobins from the nodules of non-leguminous plants [2,3], and from the roots of non-nodulating plants[4] have been recently sequenced. A signature pattern was developed that picks up the sequence of plants hemoglobins, exclusively.

Consensus pattern: [SN]-P-x-L-x(2)-H-A-x(3)-F-

[1] Powell R., Gannon F. BioEssays 9:117-121(1988).
[2] Kortt A.A., Trinick M.J., Appleby C.A. Eur. J. Biochem. 175:141-149(1988).
[3] Kortt A.A., Inglis A.S., Fleming A.I., Appleby C.A. FEBS Lett. 231:341-346(1988).
[4] Bogusz D., Appleby C.A., Landsmann J., Dennis E.S., Trinick M.J., Peacock W.J. Nature 331:178-180(1988).

262. Fructose-bisphosphate aldolase class-I active site (glycolytic_enz)

Fructose-bisphosphate aldolase [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6-bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate.There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-I aldolases [3], mainly found in higher eukaryotes, are homotetrameric enzymes which form a Schiff-base intermediate between the C-2 carbonyl group of the substrate (dihydroxyacetone phosphate)and the epsilon-amino group of a lysine residue. In vertebrates, three forms of this enzyme are found: aldolase A in muscle, aldolase B in liver and aldolase C in brain. The sequence around the lysine involved in the Schiff-base is highly conserved and can be used as a signature for this class of enzyme.

Consensus pattern: [LIVM]-x-[LIVMFYW]-E-G-x-[LS]-L-K-P-[SN] [K is involved in Schiff-base formation]-

[1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).
[2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).
[3] Freemont P.S., Dunbar B., Fothergill-Gilmore L.A. Biochem. J. 249:779-788(1988).

263. Glycosyl hydrolases family 11 active sites signatures

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family G [3] or as the glycosyl hydrolases family 11 [4,E1]. The enzymes which are currently known to belong to this family

are listed below. - Aspergillus awamori xylanase C (xynC). - Bacillus circulans, pumilus, stearothermophilus and subtilis xylanase (xynA). - Clostridium acetobutylicum xylanase (xynB). - Clostridium stercorarium xylanase A (xynA). - Fibrobacter succinogenes xylanase C (xynC) which consist of two catalytic domains that both belong to family 10. - Neocallimastix patriciarum xylanase A (xynA). - Ruminococcus flavefaciens bifunctional xylanase XYLA (xynA). This protein consists of three domains: a N-terminal xylanase catalytic domain that belongs to family 11 of glycosyl hydrolases; a central domain composed of short repeats of Gln, Asn an Trp, and a C-terminal xylanase catalytic domain that belongs to family 10 of glycosyl hydrolases. - Schizophyllum commune xylanase A. - Streptomyces lividans xylanases B (xlnB) and C (xlnC). - Trichoderma reesei xylanases I and II. Two of the conserved regions in these enzymes are centered on glutamic acidresidues which have both been shown [5], in Bacillus pumilis xylanase, to be necessary for catalytic activity. Both regions were used as signature patterns.

Consensus pattern: [PSA]-[LQ]-x-E-Y-Y-[LIVM](2)-[DE]-x-[FYWHN] [E is an active site residue]-

Consensus pattern: [LIVMF]-x(2)-E-[AG]-[YWG]-[QRFGS]-[SG]-[STAN]-G-x-[SAF] [E is an active site residue]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

[5] Ko E.P., Akatsuka H., Moriyama H., Shinmyo A., Hata Y., Katsube Y., Urabe I., Okada H. Biochem. J. 288:117-121(1992).

264. Glycosyl hydrolase family 14

This family are beta amylases.

265. Glycosyl hydrolases family 1 signatures

It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta-glucosidases (EC 3.2.1.21) from various bacteria such as Agrobacterium strain ATCC 21400, Bacillus polymyxa, and Caldocellum saccharolyticum. - Two plants (clover) beta-glucosidases (EC 3.2.1.21). - Two different beta-galactosidases (EC 3.2.1.23) from the archaebacteria Sulfolobus solfataricus (genes bgaS and lacS). - 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as Lactobacillus casei, Lactococcus lactis, and Staphylococcus aureus. - 6-phospho-beta-glucosidases (EC 3.2.1.86) from Escherichia coli (genes bglB and ascB) and from Erwinia chrysanthemi (gene arbB). - Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase). - Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62). LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is evolutionary related to the above glycosyl hydrolases. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-glucosidase from Agrobacterium, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. This region was used as a signature pattern. As a second signature pattern a conserved region was selected, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue.

Consensus pattern: [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue]
Note: this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].
Consensus pattern: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA]-

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991).
[3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990).
[4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992).
[5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).

266. Glycosyl hydrolases family 2 signatures

It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta-galactosidases (EC 3.2.1.23) from bacteria such as Escherichia coli (genes lacZ and ebgA), Clostridium acetobutylicum, Clostridium thermosulfurogenes, Klebsiella pneumoniae, Lactobacillus delbrueckii, or Streptococcus thermophilus and from the fungi Kluyveromyces lactis. - Beta-glucuronidase (EC 3.2.1.31) from Escherichia coli (gene uidA) and from mammals. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in Escherichia coli lacZ, to be the general acid/base catalyst in the active site of the enzyme. This region was used as a signature pattern. As a second signature pattern a highly conserved region was selected located some sixty residues upstream from the active site glutamate.

Consensus pattern: N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4)-

Consensus pattern: [DENQLF]-[KRVW]-N-[HRY]-[STAPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue]-

[1] Henrissat B. Biochem. J. 280:309-316(1991).

[2] Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. Gen. Microbiol. 137:369-380(1991).

[3] Gebler J.C., Aebersold R., Withers S.G. J. Biol. Chem. 267:11126-11130(1992).

267. Glycosyl hydrolases family 3 active site

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Beta glucosidases (EC 3.2.1.21) from the fungi Aspergillus wentii (A-3), Hansenula anomala, Kluyveromyces fragilis, Saccharomycopsis fibuligera, (BGL1 and BGL2), Schizophyllum commune and Trichoderma reesei (BGL1).
- Beta glucosidases from the bacteria Agrobacterium tumefaciens (Cbg1), Butyrivibrio fibrisolvens (bglA), Clostridium thermocellum (bglB),

Escherichia coli (bglX), Erwinia chrysanthemi (bgxA) and Ruminococcus albus. - Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52).

- Bacillus subtilis hypothetical protein yzbA.
- Escherichica coli hypothetical protein ycfO and HI0959, the corresponding Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus wentii beta-glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern: [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).
[3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).

268. Glycosyl hydrolases family 8 signature

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91)(exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family D [3] or as the glycosyl hydrolases family 8 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - Acetobacter xylinum endonuclease cmcAX. - Bacillus strain KSM-330 acidic endonuclease K (Endo-K). - Cellulomonas josui endoglucanase 2 (celB). - Cellulomonas uda endoglucanase. - Clostridium cellulolyticum endoglucanases C (celcCC). - Clostridium thermocellum endoglucanases A (celA). - Erwinia chrysanthemi minor endoglucanase y (celY). - Bacillus circulans beta-glucanase (EC 3.2.1.73). - Escherichia coli hypothetical protein yhjM. The most conserved region in these enzymes is a stretch of about 20 residues that contains two conserved aspartate. The first asparatate is thought [5] to act as the nucleophile in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern: A-[ST]-D-[AG]-D-x(2)-[IM]-A-x-[SA]-[LIVM]-[LIVMG]-x-A- x(3)-[FW] [The first D is an active site residue]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

[5] Alzari P.M., Souchon H., Dominguez R. Structure 4:265-275(1996).

269. Glycosyl hydrolases family 9 active sites signatures

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produce a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family E [3] or as the glycosyl hydrolases family 9 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - Butyrivibrio fibrisolvens cellodextrinase 1 (ced1). - Cellulomonas fimi endoglucanases B (cenB) and C (cenC). - Clostridium cellulolyticum endoglucanase G (celCCG). - Clostridium cellulovorans endoglucanase C (engC). - Clostridium stercoararium endoglucanase Z (avicelase I) (celZ). - Clostridium thermocellum endoglucanases D (celD), F (celF) and I (celI). - Fibrobacter succinogenes endoglucanase A (endA). - Pseudomonas fluorescens endoglucanase A (celA). - Streptomyces reticuli endoglucanase 1 (cel1). - Thermomonospora fusca endoglucanase E-4 (celD). - Dictyostelium discoideum spore germination specific endoglucanase 270-6. This slime mold enzyme may digest the spore cell wall during germination, to release the enclosed amoeba. - Endoglucanases from plants such as Avocado or French bean. In plants this enzyme may be involved the fruit ripening process. Two of the most conserved regions in these enzymes are centered on conserved residues which have been shown [5,6], in the endoglucanase D from Cellulomonas thermocellum, to be important for the catalytic activity. The first region contains an active site histidine and the second region contains two catalytically important residues: an aspartate and a glutamate. Both regions were used as signature patterns.

Consensus pattern: [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R [H is an active site residue]-

Consensus pattern: [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA] [D and E are active site residues]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

[5] Tomme P., Chauvaux S., Beguin P., Millet J., Aubert J.-P., Claeyssens M. J. Biol. Chem. 266:10313-10318(1991).

[6] Tomme P., van Beeumen J., Claeyssens M. Biochem. J. 285:319-324(1992).

270. Glyceraldehyde 3-phosphate dehydrogenase active site (gpdh)

Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH) [1] is a tetrameric NAD-binding enzyme common to both the glycolytic and gluconeogenic pathways. A cysteine in the middle of the molecule is involved in forming a covalent phosphoglycerol thioester intermediate. The sequence around this cysteine is totally conserved in eubacterial and eukaryotic GAPDHs and is also present, albeit in a variant form, in the otherwise highly divergent archaebacterial GAPDH [2].Escherichia coli D-erythrose 4-phosphate dehydrogenase (E4PDH) (gene epd orgapB) is an enzyme highly related to GAPDH [3].

Consensus pattern: [ASV]-S-C-[NT]-T-x(2)-[LIM] [C is the active site residue]-

[1] Harris J.I., Waters M. (In) The Enzymes (3rd edition) 13:1-50(1976).

[2] Fabry S., Lang J., Niermann T., Vingron M., Hensel R. Eur. J. Biochem. 179:405-413(1989).

[3] Zhao G., Pease A.J., Bharani N., Winkler M.E. J. Bacteriol. 177:2804-2812(1995).

271. Granulins signature

Granulins [1] are a family of cysteine-rich peptides of about 6 Kd which may have multiple biological activity. A precursor protein (known as acrogranin) potentially encodes seven different forms of granulin (grnA to grnG) which are probably released by post-translational proteolytic processing. A schematic representation of the structure of a granulin is shown below: xxxCxxxxxCxxxxxCCxxxxxxxxCCxxxxxxCCxxxxxCCxxxxxCxxxxxxCx *************'C': conserved cysteine probably involved in a disulfide bond.'*': position of the pattern. Granulins are evolutionary related to a PMP-D1, a peptide extracted from thepars intercerebralis of migratory locusts [2].

Consensus pattern: C-x-D-x(2)-H-C-C-P-x(4)-C [The four C's are probably involved in disulfide bonds]-

[1] Bhandari V., Palfree R.G., Bateman A. Proc. Natl. Acad. Sci. U.S.A. 89:1715-1719(1992).

[2] Nakakura N., Hietter H., van Dorsselaer A., Luu B. Eur. J. Biochem. 204:147-153(1992).

272. (HCV RdRp) Hepatitis C virus RNA dependent RNA polymerase

The RNA dependent RNA polymerase is also known as non-structural protein NS5B. NS5B is a 65 kDa protein that resembles other viral RNA polymerases. HCV replication is thought to occur in membrane bound replication complexes. These complexes transcribe the positive strand and the resulting minus strand is used as a template for the synthesis of genomic RNA. There are two viral proteins involved in the reaction, NS3 and NS5B.[1,2]

[1] Lohmann V, Korner F, Herian U, Bartenschlager R; J Virol 1997;71:8416-8428. [2] Behrens SE, Tomei L, De Francesco R; EMBO J 1996;15:12-22. [3] Ishido S, Fujita T, Hotta H; Biochem Biophys Res Commun 1998;244:35-40.

273. (HHH) Helix-hairpin-helix motif.

[1] Doherty AJ, Serpell LC, Ponting CP; Nucleic Acids Res 1996;24:2488-2497.

274. HIT family signature

Recently a family of small proteins of about 12 to 16 Kd has been described[1]. This family currently consists of: - Mammalian protein HINT (also known as Protein kinase C inhibitor 1 or PKCI- 1). HINT was incorrectly thought to be a specific inhibitor of PKC. It has been shown to bind zinc. - Fission yeast diadenosine 5',5'''-P1,P4-tetraphosphate asymmetrical hydrolase (Ap4Aase) (EC 3.6.1.17) [2] (gene aph1), which cleaves A-5'-PPPP- 5'A to yield AMP and ATP. - FHIT, a human protein whose gene is altered in different tumors and which acts [3] as a diadenosine 5',5'''-P1,P3-triphosphate hydrolase (Ap3Aase) (EC 3.6.1.29) cleaving A-5'-PPP-5'A to yield AMP and ADP. - Yeast proteins HNT1 and HNT2. - Maize zinc-binding protein ZBP14. - Escherichia coli hypothetical protein ycfF. - Haemophilus influenzae hypothetical protein HI0961. - Helicobacter pylori hypothetical protein HP0404. - Methanococcus jannaschii hypothetical protein MJ0866. - Mycobacterium leprae hypothetical protein U296A. - Synechocystis strain PCC 6803 hypothetical protein slr1234. - Caenorhabditis elegans hypothetical protein F21C3.3. - A hypothetical 13.2 Kd protein in hisE 3'region in Azospirillum brasilense. - A hypothetical 13.1 Kd protein in p37 5'region in Mycoplasma hyorhinis. - A hypothetical 12.4 Kd protein in psbAII 5'region in Synechococcus strain PCC 7942.All these proteins contains a region with three clustered histidines. This region is responsible for the designation of this family: HIT, for 'HIstidineTriad [1]. This region was originally thought to be implied in the binding of a zinc ion but was later identified [4] as part of the alpha-phosphate binding site of a nucleotide-binding domain. As a signature pattern, the region of the histidine triad was selected.

Consensus pattern: [NQA]-x(4)-[GAV]-x-[QF]-x-[LIVM]-x-H-[LIVMFYT]-H-[LIVMFT]-H-[LIVMF](2)-[PSGA]-

[1] Seraphin B. DNA Seq. 3:177-179(1992).
[2] Huang Y., Garrison P.N., Barnes L.D. Biochem. J. 312:925-932(1995).
[3] Barnes L.D., Garrison P.N., Siprashvili Z., Guranowski A., Robinson A.K., Ingram S.W., Croce C.M., Ohta M., Huebner K. Biochemistry 35:11529-11535(1996).

[4] Brenner C., Garrison P., Gilmour J., Peisach D., Ringe D., Petsko G.A., Lowenstein J.M. Nat. Struct. Biol. 4:231-238(1997).

275. Myc-type, 'helix-loop-helix' dimerization domain signature (HLH)

A number of eukaryotic proteins, which probably are sequence specific DNA-binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are refered as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also referred to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW. - The myc family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation. - Proteins involved in myogenesis (the induction of muscle cells). In mammals MyoD1 (Myf-3), myogenin (Myf-4), Myf-5, and Myf-6 (Mrf4 or herculin), in birds CMD1 (QMF-1), in Xenopus MyoD and MF25, in Caenorhabditis elegans CeMyoD, and in Drosophila nautilus (nau). - Vertebrate proteins that bind specific DNA sequences ('E boxes') in various immunoglobulin chains enhancers: E2A or ITF-1 (E12/pan-2 and E47/pan-1), ITF-2 (tcf4), TFE3, and TFEB. - Vertebrate neurogenic differentiation factor 1 that acts as differentiation factor during neurogenesis. - Vertebrate MAX protein, a transcription regulator that forms a sequence- specific DNA-binding protein complex with myc or mad. - Vertebrate Max Interacting Protein 1 (MXI1 protein) which acts as a transcriptional repressor and may antagonize myc transcriptional activity by competing for max. - Proteins of the bHLH/PAS superfamily which are transcriptional activators. In

mammals, AH receptor nuclear translocator (ARNT), single-minded homologs (SIM1 and SIM2), hypoxia-inducible factor 1 alpha (HIF1A), AH receptor (AHR), neuronal pas domain proteins (NPAS1 and NPAS2), endothelial pas domain protein 1 (EPAS1), mouse ARNT2, and human BMAL1. In drosophila, single-minded (SIM), AH receptor nuclear translocator (ARNT), trachealess protein (TRH), and similar protein (SIMA). - Mammalian transcription factors HES, which repress transcription by acting on two types of DNA sequences, the E box and the N box. - Mammalian MAD protein (max dimerizer) which acts as transcriptional repressor and may antagonize myc transcriptional activity by competing for max. - Mammalian Upstream Stimulatory Factor 1 and 2 (USF1 and USF2), which bind to a symmetrical DNA sequence that is found in a variety of viral and cellular promoters. - Human lyl-1 protein; which is involved, by chromosomal translocation, in T- cell leukemia. - Human transcription factor AP-4. - Mouse helix-loop-helix proteins MATH-1 and MATH-2 which activate E box- dependent transcription in collaboration with E47. - Mammalian stem cell protein (SCL) (also known as tal1), a protein which may play an important role in hemopoietic differentiation. SCL is involved, by chromosomal translocation, in stem-cell leukemia. - Mammalian proteins Id1 to Id4 [5]. Id (inhibitor of DNA binding) proteins lack a basic DNA-binding domain but are able to form heterodimers with other HLH proteins, thereby inhibiting binding to DNA. - Drosophila extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins. - Human Sterol Regulatory Element Binding Protein 1 (SREBP-1), a transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) found in the flanking region of the LDLR gene and in other genes. - Drosophila achaete-scute (AS-C) complex proteins T3 (l'sc), T4 (scute), T5 (achaete) and T8 (asense). The AS-C proteins are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system. - Mammalian homologs of achaete-scute proteins, the MASH-1 and MASH-2 proteins. - Drosophila atonal protein (ato) which is involved in neurogenesis. - Drosophila daughterless (da) protein, which is essential for neurogenesis and sex-determination. - Drosophila deadpan (dpn), a hairy-like protein involved in the functional differentiation of neurons. - Drosophila delilah (dei) protein, which is plays an important role in the differentiation of epidermal cells into muscle. - Drosophila hairy (h) protein, a transcriptional repressor which regulates the embryonic segmentation and adult bristle patterning. - Drosophila enhancer of split proteins E(spl), that are hairy-like proteins active during neurogenesis. also act as transcriptional repressors. - Drosophila twist (twi) protein, which is involved in the establishment of germ

layers in embryos. - Maize anthocyanin regulatory proteins R-S and LC. - Yeast centromere-binding protein 1 (CPF1 or CBF1). This protein is involved in chromosomal segregation. It binds to a highly conserved DNA sequence, found in centromers and in several promoters. - Yeast INO2 and INO4 proteins. - Yeast phosphate system positive regulatory protein PHO4 which interacts with the upstream activating sequence of several acid phosphatase genes. - Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression. - Neurospora crassa nuc-1, a protein that activates the transcription of structural genes for phosphorus acquisition. - Fission yeast protein esc1 which is involved in the sexual differentiation process. The schematic representation of the helix-loop-helix domain is shown here: xxxxxxxxxxxxxxxxxxxxxxxx-------------------xxxxxxxxxxxxxxxxxxxxxxx Amphipathic helix 1 Loop Amphipathic helix 2. The signature pattern developed to detect this domain spans completely the second amphipathic helix.

Consensus pattern: [DENSTAP]-[KTR]-[LIVMAGSNT]-{FYWCPHKR}-[LIVMT]-[LIVM]- x(2)-[STAV]-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]-{P}-{P}-[LIVMRKHQ].-

[1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).
[2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).
[3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).
[4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990).
[5] Riechmann V., van Cruechten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

276. HMG14 and HMG17 signature

High mobility group (HMG) proteins are a family of relatively low molecular weight non-histone components in chromatin. HMG14 and HMG17 [1], two related proteins of about 100 amino acid residues, bind to the inner side of the nucleosomal DNA thus altering the interaction between the DNA and the histone octamer. These two proteins may be involved in the process which maintains transcribable genes in a unique chromatin conformation. The trout nonhistone chromosomal protein H6 (histone T) also belongs to this family. As a signature pattern a conserved stretch of 10 residues located in the N-terminal section of HMG14 and HMG17 was selected.

Consensus pattern: R-R-S-A-R-L-S-A-[RK]-P-

[1] Bustin M., Reeves R. Prog. Nucleic Acid Res. Mol. Biol. 54:35-100(1996).

277. Hydroxymethylglutaryl-coenzyme A lyase active site (HMGL1)
3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4)catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In vertebrates it is a mitochondrial enyme which is involved in ketogenesis and in leucine catabolism [1]. In some bacteria, such as Pseudomonas mevalonii, it is involved in mevalonate catabolism (gene mvaB). A cysteine has been shown[2], in mvaB, to be required for the activity of the enzyme. The region around this residue is perfectly conserved and is used as a signature pattern.

Consensus pattern: S-V-A-G-L-G-G-C-P-Y [C is the active site residue]-

[1] Mitchell G.A., Robert M.-F., Hruz P.W., Wang S., Fontaine G., Behnke C.E., Mende-Mueller L.M., Schappert K., Lee C., Gibson K.M., Miziorko H.M. J. Biol. Chem. 268:4376-4381(1993).
[2] Hruz P.W., Narasimhan C., Miziorko H.M. Biochemistry 31:6842-6847(1992).

Alpha-isopropylmalate and homocitrate synthases signatures (HMGL2)
The following enzymes have been shown [1] to be functionally as well as evolutionary related: - Alpha-isopropylmalate synthase (EC 4.1.3.12) which catalyzes the first step in the biosynthesis of leucine, the condensation of acetyl-CoA and alpha- ketoisovalerate to form 2-isopropylmalate synthase. - Homocitrate synthase (EC 4.1.3.21) (gene nifV) which is involved in the biosynthesis of the iron-molybdenum cofactor of nitrogenase and catalyzes the condensation of acetyl-CoA and alpha-ketoglutarate into homocitrate. - Soybean late nodulin 56. - Methanococcus jannaschii hypothetical proteins MJ0503, MJ1195 and MJ1392. Two conserved regions were selected as signature patterns for these enzymes. The first region is located in the N-terminal section while the second region is located in the central section and contains two conserved histidine residues which could be implicated in the catalytic mechanism.

Consensus pattern: L-R-[DE]-G-x-Q-x(10)-K-

Consensus pattern: [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI]-

[1] Wang S.-Z., Dean D.R., Chen J.-S., Johnson J.L. J. Bacteriol. 173:3041-3046(1991).

278. (HMG C0A synt) Hydroxymethylglutaryl-coenzyme A synthase active site

Hydroxymethylglutaryl-coenzyme A synthase (EC 4.1.3.5) (HMG-CoA synthase) catalyzes the condensation of acetyl-CoA with acetoacetyl-CoA to produce HMG- CoA and CoA [1].In vertebrates there are two isozymes located in different subcellular compartments: a cytosolic form which is the starting point of the mevalonate pathway which leads to cholesterol and other sterolic and isoprenoid compounds and a mitochondrial form responsible for ketone body biosynthesis. HMG-CoA is also found in other eukaryotes such as insect, plants and fungi. A cysteine is known to act as the catalytic nucleophile in the first step of the reaction, the acetylation of the enzyme by acetyl-CoA. The conserved region was used around this active site residue as a signature pattern.

Consensus pattern: N-x-[DN]-[IV]-E-G-[IV]-D-x(2)-N-A-C-[FY]-x-G [C is the active site residue]-

[1] Rokosz L.L., Boulton D.A., Butkiewicz E.A., Sanyal G., Cueto M.A., Lachance P.A., Hermes J.D. Arch. Biochem. Biophys. 312:1-13(1994).

279. HMG (high mobility group) box

280. HSF-type DNA-binding domain signature

Heat shock factor (HSF) is a DNA-binding protein that specifically binds heat shock promoter elements (HSE). HSE is a palindromic element rich with repetitive purine and pyrimidine motifs: 5'-nGAAnnTTCnnGAAnnTTCn-3'. HSF is expressed at normal temperatures but is activated by heat shock or chemical stressors [1,2]. The sequences of HSF from various species show extensive similarity in a region of about 90 amino acids, which has been shown [3] to bind DNA. Some other proteins also contain a HSF domain, these are:

- Yeast SFL1, a protein involved in cell surface assembly and regulation of the gene related to flocculation (asexual cell aggregation) [4]. - Yeast transcription factor SKN7 (or BRY1 or POS9), which binds to the promoter elements SCB and MCB essential for the control of G1 cyclins expression [5]. - Yeast MGA1. - Yeast hypothetical protein YJR147w. A pattern from the most conserved part of the HSF DNA-binding domain was derived, its central region.

Consensus pattern: L-x(3)-[FY]-K-H-x-N-x-[STAN]-S-F-[LIVM]-R-Q-L-[NH]-x-Y-x-[FYW]-[RKH]-K-[LIVM]-

[1] Sorger P.K. Cell 65:363-366(1991).
[2] Mager W.H., Moradas Ferreira P. Biochem. J. 290:1-13(1993).
[3] Vuister G.W., Kim S.-J., Orosz A., Marquardt J., Wu C., Bax A. Nat. Struct. Biol. 1:605-613(1994).
[4] Fujita A., Kikuchi Y., Kuhara S., Misumi Y., Matsumoto S., Kobayashi H. Gene 85:321-328(1989).
[5] Morgan B.A., Bouquin N., Merrill G.F., Johnston L.H. EMBO J. 14:5679-5689(1995).

281. Heat shock hsp20 proteins family profile

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by inducing the synthesis of proteins collectively known as heat-shock proteins (hsp) [1]. Amongst them is a family of proteins with an average molecular weight of 20 Kd, known as the hsp20 proteins [2 to 5]. These seem to act as chaperones that can protect other proteins against heat-induced denaturation and aggregation. Hsp20 proteins seem to form large heterooligomeric aggregates; their family is currently composed of the following members: - Vertebrate heat shock protein hsp27 (hsp25), induced by a variety of environmental stresses. - Drosophila heat shock proteins hsp22, hsp23, hsp26, hsp27, hsp67BA and BC. - Caenorhabditis elegans hsp16 multigene family. - Fungal HSP26 (budding yeast) and hsp30 (Neurospora crassa and Aspergillus Nidulans). - Plant small hsp's. Plants have four classes of hsp20: classes I and II which are cytoplasmic, class III which is chloroplastic and class IV which is found in the endomembrane. - Alpha-crystallin A and B chains. Alpha-crystallin is an abundant constituent of the eye lens of most vertebrate species. Its main function appears to be to maintain the correct refractive index of the lens. It is also found in other tissues where it seems to act as a chaperone [6]. - Schistosoma mansoni major egg antigen p40.

Structurally, p40 is built of two tandem hsp20 domains. - A variety of prokaryotic proteins: ibpA and ibpB from Escherichia coli, hsp18 from Clostridium acetobutylicum, spore protein SP21 (hspA) from Stigmatella aurantiaca, Mycobacterium leprae 18 Kd antigen and Mycobacterium tuberculosis 14 Kd antigen. - Methanococcus jannaschii hypothetical protein MJ0285.Structurally, this family is characterized by the presence of a conserved C-terminal domain of about 100 residues. The profile developed to detect members of the hsp20 family is based on an alignment of this domain.

-Sequences known to belong to this class detected by the profile: ALL.

[1] Lindquist S., Craig E.A. Annu. Rev. Genet. 22:631-677(1988).[2] de Jong W.W., Leunissen J.A.M., Voorter C.E.M. Mol. Biol. Evol. 10:103-126(1993).[3] Caspers G.J., Leunissen J.A.M., de Jong W.W. J. Mol. Evol. 40:238-248(1995).[4] Jaenicke R., Creighton T.E. Curr. Biol. 3:234-235(1993).[5] Jakob U., Buchner J. Trends Biochem. Sci. 19:205-211(1994).[6] Groenen P.J.T.A., Merck K.B., de Jong W.W., Bloemendal H. Eur. J. Biochem. 225:1-9(1994).

282. Heat shock hsp70 proteins family signatures

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by the induction of the synthesis of proteins collectively known as heat-shock proteins (hsp) [1]. Amongst them is a family of proteins with an average molecular weight of 70 Kd, known as the hsp70proteins [2,3,4]. In most species, there are many proteins that belong to the hsp70 family. Some of them are expressed under unstressed conditions. Hsp70proteins can be found in different cellular compartments (nuclear, cytosolic, mitochondrial, endoplasmic reticulum, etc.). Some of the hsp70 family proteinsare listed below: - In Escherichia coli and other bacteria, the main hsp70 protein is known as the dnaK protein. A second protein, hscA, has been recently discovered. dnaK is also found in the chloroplast genome of red algae. - In yeast, at least ten hsp70 proteins are known to exist: SSA1 to SSA4, SSB1, SSB2, SSC1, SSD1 (KAR2), SSE1 (MSI3) and SSE2. - In Drosophila, there are at least eight different hsp70 proteins: HSP70, HSP68, and HSC-1 to HSC-6. - In mammals, there are at least eight different proteins: HSPA1 to HSPA6, HSC70, and GRP78 (also known as the immunoglobulin heavy chain binding protein (BiP)). - In the sugar beet yellow virus (SBYV), a hsp70 homolog has been shown [5] to exist. - In archaebacteria, hsp70 proteins are also present [6].All proteins belonging to the hsp70 family bind ATP. A variety of functions has been postulated for hsp70 proteins. It now appears [7] that some hsp70proteins

play an important role in the transport of proteins across membranes. They also seem to be involved in protein folding and in the assembly/disassembly of protein complexes [8]. Three signature patterns for the hsp70 family of proteins were derived; the first centered on a conserved pentapeptide found in the N-terminal section of these proteins; the two others on conserved regions located in the central part of the sequence.

Consensus pattern: [IV]-D-L-G-T-[ST]-x-[SC] -

Consensus pattern: [LIVMF]-[LIVMFY]-[DN]-[LIVMFS]-G-[GSH]-[GS]-[AST]-x(3)- [ST]-[LIVM]-[LIVMFC]-

Consensus pattern: [LIVMY]-x-[LIVMF]-x-G-G-x-[ST]-x-[LIVM]-P-x-[LIVM]-x-[DEQKRSTA]-

[1] Lindquist S., Craig E.A. Annu. Rev. Genet. 22:631-677(1988).

[2] Pelham H.R.B. Cell 46:959-961(1986).

[3] Pelham H.R.B. Nature 332:776-77(1988).[4] Craig E.A. BioEssays 11:48-52(1989).

[5] Agranovsky A.A., Boyko V.P., Karasev A.V., Koonin E.V., Dolja V.V. J. Mol. Biol. 217:603-610(1991).

[6] Gupta R.S., Singh B. J. Bacteriol. 174:4594-4605(1992).

[7] Deshaies R.J., Koch B.D., Schekmam R. Trends Biochem. Sci. 13:384-388(1988).

[8] Craig E.A., Gross C.A. Trends Biochem. Sci. 16:135-140(1991).

283. Heat shock hsp90 proteins family signature

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by the induction of the synthesis of proteins collectively known as heat-shock proteins (hsp) [1]. Amongst them is a family of proteins, with an average molecular weight of 90 Kd, known as the hsp90proteins. Proteins known to belong to this family are: - Escherichia coli and other bacteria heat shock protein c62.5 (gene htpG). - Vertebrate hsp 90-alpha (hsp 86) and hsp 90-beta (hsp 84). - Drosophila hsp 82 (hsp 83). - Trypanosoma cruzi hsp 85. - Plants Hsp82 or Hsp83. - Yeast and other fungi HSC82, and HSP82. - The endoplasmic reticulum protein 'endoplasmin' (also known as Erp99 in mouse, GRP94 in hamster, and hsp 108 in chicken).The exact function of hsp90 proteins is not yet known. In higher eukaryotes, hsp90 has been found associated with steroid hormone receptors, with tyrosine kinase oncogene products of several retroviruses, with eIF2alpha kinase, and with actin and tubulin. Hsp90 are

probable chaperonins that possess ATPase activity [2,3].As a signature pattern for the hsp90 family of proteins, a highly conserved region found in the N-terminal part of these proteins was selected.

Consensus pattern: Y-x-[NQH]-K-[DE]-[IVA]-F-L-R-[ED] -

[1] Lindquist S., Craig E.A. Annu. Rev. Genet. 22:631-677(1988).
[2] Nadeau K., Das A., Walsh C.T. J. Biol. Chem. 268:1479-1487(1993).
[3] Jakob U., Buchner J. Trends Biochem. Sci. 19:205-211(1994).

284. Helix-turn-helix (HTH3)

This large family of DNA binding helix-turn helix proteins includes Cro Swiss:P03036 and CI Swiss:P03034.

285. Heme oxygenase signature

Heme oxygenase (EC 1.14.99.3) (HO) [1] is the microsomal enzyme that, in animals, carries out the oxidation of heme, it cleaves the heme ring at the alpha methene bridge to form biliverdin and carbon monoxide. Biliverdin is subsequently converted to bilirubin by biliverdin reductase. In mammals there are three isozymes of heme oxygenase: HO-1 to HO-3. The first two isozymes differ in their tissue expression and their inducibility: HO-1 is highly inducible by its substrate heme and by various non-heme substances, while HO-2 is non-inducible. It has been suggested [2] that HO-2 could be implicated in the production of carbon monoxide in the brain where it is said to act as a neurotransmitter.In the genome of the chloroplast of red algae as well as in cyanobacteria, there is a heme oxygenase (gene pbsA) that is the key enzyme in the synthesis of the chromophoric part of the photosynthetic antennae [3]. An heme oxygenase is also present in the bacteria Corynebacterium diphtheriae (gene hmuO), where it is involved in the acquisition of iron from the host heme [4].There is, in the central section of these enzymes, a well conserved region centered on a histidine residue which is proposed to play a key role in binding the substrate heme at the active center of the enzyme. This region was used as a signature pattern.

Consensus pattern: L-[IV]-A-H-[STACH]-Y-[STV]-[RT]-Y-[LIVM]-G [H binds the heme] -

[1] Maines M.D. FASEB J. 2:2557-2568(1988).
[2] Barinaga M. Science 259:309-309(1993).
[3] Richaud C., Zabulon G. Proc. Natl. Acad. Sci. U.S.A. 94:11736-11741(1997).
[4] Schmitt M.P. J. Bacteriol. 179:838-845(1997).

286. Hepatitis core antigen.

The core antigen of hepatitis viruses possesses a carboxyl terminus rich in arginine. On this basis it was predicted that the core antigen would bind DNA [1]. There is some experimental evidence to support this [2].

[1] Pasek M, Goto T, Gilbert W, Zink B, Schaller H, Mckay P, Leadbetter G, Murray K; Nature 1979;282:575-579. [2] Gallina A, Bonelli F, Zentilin L, Rindi G, Muttini M, Milanesi G; J Virol 1989;63:4645-4652.

287. Histidine biosynthesis protein

Proteins involved in steps 4 and 6 of the histidine biosynthesis pathway are contained in this family. Histidine is formed by several complex and distinct biochemical reactions catalysed by eight enzymes. The enzymes in this Pfam entry are called His6 and His7 in eukaryotes and HisA and HisF in prokaryotes.

[1] Fani R, Tamburini E, Mori E, Lazcano A, Lio P, Barberio C, Casalone E, Cavalieri D, Perito B, Polsinelli M, Gene 1997;197:9-17. [2] Fani R, Lio P, Chiarelli I, Bazzicalupo M, J Mol Evol 1994;38:489-495.

288. Histone deacetylase family

Histones can be reversibly acetylated on several lysine residues. Regulation of transcription is caused in part by this mechanism. Histone deacetylases catalyse the removal of the acetyl group. Histone deacetylases are related to other proteins [1].

Leipe DD, Landsman D, Nucleic Acids Res 1997;25:3693-3697.

289. Histidinol dehydrogenase signature

Histidinol dehydrogenase (EC 1.1.1.23) (HDH) catalyzes the terminal step in the biosynthesis of histidine in bacteria, fungi, and plants, the four-electron oxidation of L-histidinol to histidine.In bacteria HDH is a single chain polypeptide; in fungi it is the C-terminal domain of a multifunctional enzyme which catalyzes three different steps of histidine biosynthesis; and in plants it is expressed as nuclear encoded protein precursor which is exported to the chloroplast [1].As a signature pattern a highly conserved region located in the central part of HDH was selected. This region does not correspond to the part of the enzyme that, in most, but not all HDH sequences contains a cysteine residue which, in Salmonella typhimurium, has been said [2] to be important for the catalytic activity of the enzyme.

Consensus pattern: I-D-x(2)-A-G-P-[ST]-E-[LIVS]-[LIVMA](3)-[AC]-x(3)-A-x(4)- [LIVM]-[AV]-[SACL]-[DE]-[LIVMFC]-[LIVM]-[SA]-x(2)-E-H-

[1] Nagai A., Ward E., Beck J., Tada S., Chang J.-Y., Scheidegger A., Ryals J. Proc. Natl. Acad. Sci. U.S.A. 88:4133-4137(1991).

[2] Grubmeyer C.T., Gray W.R. Biochemistry 25:4778-4784(1986).

290. Homoserine dehydrogenase signature

Homoserine dehydrogenase (EC 1.1.1.3) (HDh) [1,2] catalyzes NAD-dependent reduction of aspartate beta-semialdehyde into homoserine. This reaction is the third step in a pathway leading from aspartate to homoserine. The latter participates in the biosynthesis of threonine and then isoleucine as well as in that of methionine. HDh is found either as a single chain protein as in some bacteria and yeast, or as a bifunctional enzyme consisting of an N-terminal aspartokinase domain and a C-terminal HDh domain as in bacteria such as Escherichia coli and in plants. As a signature pattern, the best conserved region of Hdh has been selected. This is a segment of 23 to 24 residues located in the central section and that contains two conserved aspartate residues.

Consensus pattern: A-x(3)-G-[LIVMFY]-[STAG]-x(2,3)-[DNS]-P-x(2)-D-[LIVM]-x-G- x-D-x(3)-K-

[1] Thomas D., Barbey R., Surdin-Kerjan Y. FEBS Lett. 323:289-293(1993).
[2] Cami B., Clepet C., Patte J.-C. Biochimie 75:487-495(1993).

291. haloacid dehalogenase-like hydrolase

This family is structurally different from the alpha/ beta hydrolase family (abhydrolase). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment, between residues 16 and 96 of Swiss:P24069. The rest of the fold is composed of the core alpha/beta domain.

[1] Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara T, Esaki N, Soda K, J Biol Chem 1996; 271:20322-20330.

292. DEAD and DEAH box families ATP-dependent helicases signatures (helicase_C)

A number of eukaryotic and prokaryotic proteins have been characterized [1,2,3] on the basis of their structural similarity. They all seem to be involved in ATP-dependent, nucleic-acid unwinding. Proteins currently known to belong to this family are: - Initiation factor eIF-4A. Found in eukaryotes, this protein is a subunit of a high molecular weight complex involved in 5'cap recognition and the binding of mRNA to ribosomes. It is an ATP-dependent RNA-helicase. - PRP5 and PRP28. These yeast proteins are involved in various ATP-requiring steps of the pre-mRNA splicing process. - Pl10, a mouse protein expressed specifically during spermatogenesis. - An3, a Xenopus putative RNA helicase, closely related to Pl10. - SPP81/DED1 and DBP1, two yeast proteins probably involved in pre-mRNA splicing and related to Pl10. - Caenorhabditis elegans helicase glh-1. - MSS116, a yeast protein required for mitochondrial splicing. - SPB4, a yeast protein involved in the maturation of 25S ribosomal RNA. - p68, a human nuclear antigen. p68 has ATPase and DNA-helicase activities in vitro. It is involved in cell growth and division. - Rm62 (p62), a Drosophila putative RNA helicase related to p68. - DBP2, a yeast protein related to p68. - DHH1, a yeast protein. - DRS1, a yeast protein involved in ribosome assembly. - MAK5, a yeast protein involved in maintenance of dsRNA killer plasmid. - ROK1, a yeast protein. - ste13, a fission

yeast protein. - Vasa, a Drosophila protein important for oocyte formation and specification of embryonic posterior structures. - Me31B, a Drosophila maternally expressed protein of unknown function. - dbpA, an Escherichia coli putative RNA helicase. - deaD, an Escherichia coli putative RNA helicase which can suppress a mutation in the rpsB gene for ribosomal protein S2. - rhlB, an Escherichia coli putative RNA helicase. - rhlE, an Escherichia coli putative RNA helicase. - srmB, an Escherichia coli protein that shows RNA-dependent ATPase activity. It probably interacts with 23S ribosomal RNA. - Caenorhabditis elegans hypothetical proteins T26G10.1, ZK512.2 and ZK686.2. - Yeast hypothetical protein YHR065c. - Yeast hypothetical protein YHR169w. - Fission yeast hypothetical protein SpAC31A2.07c. - Bacillus subtilis hypothetical protein yxiN. All these proteins share a number of conserved sequence motifs. Some of them are specific to this family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases `superfamily' [4,E1]. One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins [3,5,6,E1]. Proteins currently known to belong to this subfamily are: - PRP2, PRP16, PRP22 and PRP43. These yeast proteins are all involved in various ATP-requiring steps of the pre-mRNA splicing process. - Fission yeast prh1, which my be involved in pre-mRNA splicing. - Male-less (mle), a Drosophila protein required in males, for dosage compensation of X chromosome linked genes. - RAD3 from yeast. RAD3 is a DNA helicase involved in excision repair of DNA damaged by UV light, bulky adducts or cross-linking agents. Fission yeast rad15 (rhp3) and mammalian DNA excision repair protein XPD (ERCC-2) are the homologs of RAD3. - Yeast CHL1 (or CTF1), which is important for chromosome transmission and normal cell cycle progression in G(2)/M. - Yeast TPS1. - Yeast hypothetical protein YKL078w. - Caenorhabditis elegans hypothetical proteins C06E1.10 and K03H1.2. - Poxviruses' early transcription factor 70 Kd subunit which acts with RNA polymerase to initiate transcription from early gene promoters. - I8, a putative vaccinia virus helicase. - hrpA, an Escherichia coli putative RNA helicase. Signature patterns were developed for both subfamilies.

Consensus pattern: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]-
Consensus pattern: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR] -
Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see the relevant entry <PDOC00017

[1] Schmid S.R., Linder P. Mol. Microbiol. 6:283-292(1992).

[2] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[3] Wassarman D.A., Steitz J.A. Nature 349:463-464(1991).

[4] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[5] Harosh I., Deschavanne P. Nucleic Acids Res. 19:6331-6331(1991).

[6] Koonin E.V., Senkevich T.G. J. Gen. Virol. 73:989-993(1992).

293. Heme-binding domain in cytochrome b5 and oxidoreductases (heme_1)

Cytochrome b5 is a membrane-bound hemo protein which acts as an electron carrier for several membrane-bound oxygenases [1]. There are two homologous forms of b5, one found in microsomes and one found in the outer membrane of mitochondria. Two conserved histidine residues serve as axial ligands for the heme group. The structure of a number of oxidoreductases consists of the juxtaposition of a heme-binding domain homologous to that of b5 and either a flavodehydrogenase or a molybdopterin domain. These enzymes are:

- Lactate dehydrogenase (EC 1.1.2.3) [2], an enzyme that consists of a flavodehydrogenase domain and a heme-binding domain called cytochrome b2.
- Nitrate reductase (EC 1.6.6.1), a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria [3,4]. Consists of a molybdopterin domain (see <PDOC00484>), a heme-binding domain called cytochrome b557, as well as a cytochrome reductase domain.
- Sulfite oxidase (EC 1.8.3.1) [5], which catalyzes the terminal reaction in the oxidative degradation of sulfur-containing amino acids. Also consists of a molybdopterin domain and a heme-binding domain.

This family of proteins also includes:

- TU-36B, a Drosophila muscle protein of unknown function [6].
- Fission yeast hypothetical protein SpAC1F12.10c.
- Yeast hypothetical protein YMR073c.
- Yeast hypothetical protein YMR272c.

A segment was used which includes the first of the two histidine heme ligands, as a signature pattern for the heme-binding domain of cytochrome b5 family.

Consensus pattern: [FY]-[LIVMK]-x(2)-H-P-[GA]-G [H is a heme axial ligand]-

[1] Ozols J. Biochim. Biophys. Acta 997:121-130(1989).

[2] Guiard B. EMBO J. 4:3265-3272(1985).

[3] Calza R., Huttner E., Vincentz M., Rouze P., Galangau F., Vaucheret H., Cherel I., Meyer C., Kronenberger J., Caboche M. Mol. Gen. Genet. 209:552-562(1987).

[4] Crawford N.M., Smith M., Bellissimo D., Davis R.W. Proc. Natl. Acad. Sci. U.S.A. 85:5006-5010(1988).

[5] Guiard B., Lederer F. Eur. J. Biochem. 100:441-453(1979).

[6] Levin R.J., Boychuk P.L., Croniger C.M., Kazzaz J.A., Rozek C.E. Nucleic Acids Res. 17:6349-6367(1989).

294. Hexapeptide-repeat containing-transferases signature

On the basis of sequence similarity, a number of transferases have been proposed [1,2,3,4] to belong to a single family. These proteins are: - Serine acetyltransferase (EC 2.3.1.30) (SAT) (gene cysE), an enzyme involved in cysteine biosynthesis. - Azotobacter chroococcum nitrogen fixation protein nifP. NifP is most probably a SAT involved in the optimization of nitrogenase activity. - Escherichia coli thiogalactoside acetyltransferase (EC 2.3.1.18) (gene lacA), an enzyme involved in the biosynthesis of lactose. - UDP-N-acetylglucosamine acyltransferase (EC 2.3.1.129) (gene lpxA), an enzyme involved in the biosynthesis of lipid A, a phosphorylated glycolipid that anchors the lipopolysaccharide to the outer membrane of the cell. - UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (EC 2.3.1.-) (gene lpxD or firA), which is also involved in the biosynthesis of lipid A. - Chloramphenicol acetyltransferase (CAT) (EC 2.3.1.28) from Agrobacterium tumefaciens, Bacillus sphaericus, Escherichia coli plasmid IncFII NR79, Pseudomonas aeruginosa, Staphylococcus aureus plasmid pIP630. These CAT are not evolutionary related to the main family of CAT (see <PDOC00093>). - Rhizobium nodulation protein nodL. NodL is an acetyltransferase involved in the O-acetylation of Nod factors. - Bacterial maltose O-acetyltransferase (EC 2.3.1.79). - Bacterial tetrahydrodipicolinate N-succinyltransferase (EC 2.3.1.117) (gene dapD) which catalyzes the fourth step in the biosynthesis of diaminopimelate and lysine from aspartate semialdehyde. - Bacterial N-acetylglucosamine-1-phosphate uridyltransferase (EC 2.7.7.23) (gene glmU or gcaD or tms), an enzyme involved in peptidoglycan and lipopolysaccharide biosynthesis. - Staphylococcus aureus protein capG which is involved in

biosynthesis of type 1 capsular polysaccharide. - Yeast hypothetical protein YJL218w, which is highly similar to Escherichia coli lacA. - Fission yeast hypothetical protein SpAC18B11.09c. - Methanococcus jannaschii hypothetical protein MJ1064.These proteins have been shown [3,4] to contain a repeat structure composed of tandem repeats of a [LIV]-G-x(4) hexapeptide which, in the tertiary structure of lpxA [5], has been shown to form a left-handed parallel beta helix. Our signature pattern is based on a fourfold repeat of this hexapeptide.

Consensus pattern: [LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIVAC]-x-[LIV]- [GAED]-x(2)-[STAVR]-x-[LIV]-[GAED]-x(2)-[STAV]-x-[LIV]- x(3)-[LIV]-

[1] Downie J.A. Mol. Microbiol. 3:1649-1651(1989).
[2] Parent R., Roy P.H. J. Bacteriol. 174:2891-2897(1992).
[3] Vaara M. FEMS Microbiol. Lett. 97:249-254(1992).
[4] Vuorio R., Haerkonen T., Tolvanen M., Vaara M. FEBS Lett. 337:289-292(1994).
[5] Raetz C.R.H., Roderick S.L. Science 270:997-1000(1995).

295. Hexokinases signature. Hexokinase (EC 2.7.1.1) [1,2] is an important glycolytic enzyme that catalyzes the phosphorylation of keto- and aldohexoses (e.g. glucose, mannose and fructose) using MgATP as the phosphoryl donor. In vertebrates there are four major isoenzymes, commonly referred as types I,II, III and IV. Type IV hexokinase, which is often incorrectly designated glucokinase [3], is only expressed in liver and pancreatic beta-cells and plays an important role in modulating insulin secretion; it is a protein of a molecular mass of about 50 Kd. Hexokinases of types I to III, which have low Km values for glucose, have a molecular mass of about 100 Kd. Structurally they consist of a very small N-terminal hydrophobic membrane-binding domain followed by two highly similar domains of 450 residues. The first domain has lost its catalytic activity and has evolved into a regulatory domain. In yeast there are three different isozymes: hexokinase PI (gene HXK1), PII(gene HXKB), and glucokinase (gene GLK1). All three proteins have a molecular mass of about 50 Kd. All these enzymes contain one (or two in the case of types I to III isozymes)strongly conserved region which has been shown [4] to be involved in substrate binding. A pattern from that region has been derived

Consensus pattern: [LIVM]-G-F-[TN]-F-S-[FY]-P-x(5)-[LIVM]-[DNST]-x(3)-[LIVM]- x(2)-W-T-K-x-[LF]-

[1] Middleton R.J. Biochem. Soc. Trans. 18:180-183(1990).[2] Griffin L.D., Gelb B.D., Wheeler D.A., Davison D., Adams V., McCabe E.R. Genomics 11:1014-1024(1991).[3] Cornish-Bowden A., Luz Cardenas M. Trends Biochem. Sci. 16:281-282(1991).[4] Schirch D.M., Wilson J.E. Arch. Biochem. Biophys. 254:385-396(1987).

296. Histone H2A signature (his1)

Histone H2A is one of the four histones, along with H2B, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2Asequences [1,2,E1] as a signature pattern, a conserved region in the N-terminal part of H2A. This region is conserved both in classical S-phase regulated H2A's and in variant histone H2A's which are synthesized throughout the cell cycle.

Consensus pattern: [AC]-G-L-x-F-P-V-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).
[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

Histone H4 signature (his2)

Histone H4 is one of the four histones, along with H2A, H2B and H3, which forms the eukaryotic nucleosome core. Along with H3, it plays a central role in nucleosome formation. The sequence of histone H4 has remained almost invariant in more then 2 billion years of evolution [1,E1]. The region used as a signature pattern is a pentapeptide found in positions 14 to 18 of all H4sequences. It contains a lysine residue which is often acetylated [2] and a histidine residue which is implicated in DNA-binding [3].

Consensus pattern: G-A-K-R-H-

[1] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).
[2] Doenecke D., Gallwitz D. Mol. Cell. Biochem. 44:113-128(1982).
[3] Ebralidse K.K., Grachev S.A., Mirzabekov A.D. Nature 331:365-367(1988).

Histone H3 signatures (his3)

Histone H3 is one of the four histones, along with H2A, H2B and H4, which forms the eukaryotic nucleosome core. It is a highly conserved protein of 135 amino acid residues [1,2,E1].The following proteins have been found to contain a C-terminal H3-like domain: - Mammalian centromeric protein CENP-A [3]. Could act as a core histone necessary for the assembly of centromeres. - Yeast chromatin-associated protein CSE4 [4]. - Caenorhabditis elegans chromosome III encodes two highly related proteins (F54C8.2 and F58A4.3) whose C-terminal section is evolutionary related to the last 100 residues of H3. The function of these proteins is not yet known. Two signature patterns were developed, The first one corresponds to a perfectly conserved heptapeptide in the N-terminal part of H3. The second one is derived from a conserved region in the central section of H3.

Consensus pattern: K-A-P-R-K-Q-L-

Consensus pattern: P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV]-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).

[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

[3] Sullivan K.F., Hechenberger M., Masri K. J. Cell Biol. 127:581-592(1994).

[4] Stoler S., Keith K.C., Curnick K.E., Fitzgerald-Hayes M. Genes Dev. 9:573-586(1995).

Histone H2B signature (his4)

Histone H2B is one of the four histones, along with H2A, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2Bsequences [1,2,E1], a conserved region was selected in the C-terminal part ofH2B.

Consensus pattern: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L- x-[KR]-H-A-[LIVM]-[STA]-E-G-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).

[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

297. 'Homeobox' domain signature and profile (home1)

The 'homeobox' is a protein domain of 60 amino acids [1 to 5,E1] first identified in a number of Drosophila homeotic and segmentation proteins. It has since been found to be extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Some of the proteins which contain a homeobox domain play an important role in development. Most of these proteins are known to be sequence specific DNA-binding transcription factors. The homeobox domain has also been found to be very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion. A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHtttHHHHHHHHHxxxxxxxxxx | | | | | | | 1 10 20 30 40 50 60 The pattern to detect homeobox sequences that was developed is 24 residues long and spans positions 34 to 57 of the homeobox domain.

Consensus pattern: [LIVMFYG]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-[LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)- [RKNAIMW] -

[1] Gehring W.J. (In) Guidebook to the homebox genes, Duboule D., Ed., pp1-10, Oxford University Press, Oxford, (1994).

[2] Buerglin T.R. (In) Guidebook to the homebox genes, Duboule D., Ed., pp25-72, Oxford University Press, Oxford, (1994).

[3] Gehring W.J. Trends Biochem. Sci. 17:277-280(1992).

[4] Gehring W.J., Hiromi Y. Annu. Rev. Genet. 20:147-173(1986).

[5] Schofield P.N. Trends Neurosci. 10:3-6(1987).

'Homeobox' antennapedia-type protein signature (home2)

The homeotic Hox proteins are sequence-specific transcription factors. They are part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior (A-P) axis [1]. The hox proteins contain a 'homeobox' domain. In Drosophila and other insects, there are eight different Hox genes that are encoded in two gene complexes, ANT-C and BX-C. In vertebrates there are 38 genes organized in four complexes. In six of the eight Drosophila Hox genes the homeobox domain is highly similar and a conserved hexapeptide is found five to sixteen amino acids upstream of the homeobox

domain. The six Drosophila proteins that belong to this group are antennapedia (Antp), abdominal-A (abd-A), deformed (Dfd), proboscipedia (pb),sex combs reduced (scr) and ultrabithorax (ubx) and are collectively known as the 'antennapedia' subfamily. In vertebrates the corresponding Hox genes are known [2] as Hox-A2, A3, A4,A5, A6, A7, Hox-B1, B2, B3, B4, B5, B6, B7, B8, Hox-C4, C5, C6, C8, Hox-D1,D3, D4 and D8.Caenorhabditis elegans lin-39 and mab-5 are also members of the 'antennapedia' subfamily. As a signature pattern for this subfamily of homeobox proteins, the conserved hexapeptide was used.

Consensus pattern: [LIVMFE]-[FY]-P-W-M-[KRQTA]-

[1] McGinnis W., Krumlauf R. Cell 68:283-302(1992).
[2] Scott M.P. Cell 71:551-553(1992).

'Homeobox' engrailed-type protein signature (home3)

Most proteins which contain a 'homeobox' domain can be classified [1,2], on the basis of their sequence characteristics, in three subfamilies: engrailed, antennapedia and paired. Proteins currently known to belong to the engrailed subfamily are: - Drosophila segmentation polarity protein engrailed (en) which specifies the body segmentation pattern and is required for the development of the central nervous system. - Drosophila invected protein (inv). - Silk moth proteins engrailed and invected, which may be involved in the compartmentalization of the silk gland. - Honeybee E30 and E60. - Grasshopper (Schistocerca americana) G-En. - Mammalian and birds En-1 and En-2. - Zebrafish Eng-1, -2 and -3. - Sea urchin (Tripneusteas gratilla) SU-HB-en. - Leech (Helobdella triserialis) Ht-En. - Caenorhabditis elegans ceh-16.Engrailed homeobox proteins are characterized by the presence of a conserved region of some 20 amino-acid residues located at the C-terminal of the 'homeobox' domain. As a signature pattern for this subfamily of proteins, a stretch of eight perfectly conserved residues in this region was used.

Consensus pattern: L-M-A-[EQ]-G-L-Y-N-

[1] Scott M.P., Tamkun J.W., Hartzell G.W. III Biochim. Biophys. Acta 989:25-48(1989).
[2] Gehring W.J. Science 236:1245-1252(1987).

298. Isocitrate lyase signature (ICL)

Isocitrate lyase (EC 4.1.3.1) [1,2] is an enzyme that catalyzes the conversion of isocitrate to succinate and glyoxylate. This is the first step in the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. A cysteine, a histidine and a glutamate or aspartate have been found to be important for the enzyme's catalytic activity. Only one cysteine residue is conserved between the sequences of the fungal, plant and bacterial enzymes; it is located in the middle of a conserved hexapeptide that can be used as a signature pattern for this type of enzyme.

Consensus pattern: K-[KR]-C-G-H-[LMQ] [C is a putative active site residue]-

[1] Beeching J.R. Protein Seq. Data Anal. 2:463-466(1989).

[2] Atomi H., Ueda M., Hikida M., Hishida T., Teranishi Y., Tanaka A. J. Biochem. 107:262-266(1990).

299. Initiation factor 2 subunit

This family includes initiation factor 2B alpha, beta and delta subunits from eukaryotes, related proteins from archaebacteria and IF-2 from prokaryotes. Initiation factor 2 binds to Met-tRNA, GTP and the small ribosomal subunit.

[1] Kyrpides NC, Woese CR, Proc Natl Acad Sci U S A 1998;95:3726-3730.

300. Initiation factor 3 signature

Initiation factor 3 (IF-3) (gene infC) [1] is one of the three factors required for the initiation of protein biosynthesis in bacteria. IF-3 is thought to function as a fidelity factor during the assembly of the ternary initiation complex which consist of the 30S ribosomal subunit, the initiator tRNA and the messenger RNA. IF-3 binds to the 30S ribosomal subunit; it is a basic protein of 141 to 212 residues. The chloroplast initiation factor IF-3(chl) is a protein that enhances the poly(A,U,G)-dependent binding of the initiator tRNA to chloroplast ribosomal30s subunits. In its mature form it is a protein of about 400 residues whose central section is evolutionary related to the sequence of bacterial IF-3 [2].As a signature pattern a highly conserved region was selected located in the central section of bacterial IF-3 and of IF-3(chl).

Consensus pattern: [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-[LIVMFYS]-x-[FY]-[DEQTH]-x(2)-[KRQ]-

[1] Liveris D., Schwartz J.J., Geertman R., Schwartz I. FEMS Microbiol. Lett. 112:211-216(1993).

[2] Lin Q., Ma L., Burkhart W., Spremulli L.L. J. Biol. Chem. 269:9436-9444(1994).

301. Imidazoleglycerol-phosphate dehydratase signatures (IGPD)

Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) is the enzyme that catalyzes the seventh step in the biosynthesis of histidine in bacteria, fungi and plants. In most organisms it is a monofunctional protein of about 22 to29 Kd. In some bacteria such as Escherichia coli it is the C-terminal domain of a bifunctional protein that include a histidinol-phosphatase domain [1]. Two signature patterns were developed that each include two consecutive histidine residues.

Consensus pattern: [LIVMY]-[DE]-x-H-H-x(2)-E-x(2)-[GCA]-[LIVM]-[STAC]-[LIVM]-

Consensus pattern: G-x-[DN]-x-H-H-x(2)-E-[STAGC]-x-[FY]-K -

[1] Carlomagno M.S., Chiariotti L., Alifano P., Nappo A.G., Bruni C.B. J. Mol. Biol. 203:585-606(1988).

302. Indole-3-glycerol phosphate synthase signature (IGPS)

Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS) catalyzes the fourth step in the biosynthesis of tryptophan: the ring closure of 1-(2-carboxy-phenylamino)-1-deoxyribulose into indol-3-glycerol-phosphate.In some bacteria, IGPS is a single chain enzyme. In others - such as Escherichia coli - it is the N-terminal domain of a bifunctional enzyme that also catalyzes N-(5'-phosphoribosyl)anthranilate isomerase (PRAI) activity, the third step of tryptophan biosynthesis. In fungi, IGPS is the central domain of a trifunctional enzyme that also contains a PRAI C-terminal domain and a glutamine amidotransferase N-terminal domain. The N-terminal section of IGPS contains a highly conserved region which X-ray

crystallography studies [1] have shown to be part of the active site cavity. This region was used as a signature pattern for IGPS.

Consensus pattern: [LIVMFY]-[LIVMC]-x-E-[LIVMFYC]-K-[KRSP]-[STAK]-S-P-[ST]-x(3)-[LIVMFYST]-

[1] Wilmanns M., Priestle J.P., Niermann T., Jansonius J.N. J. Mol. Biol. 223:477-507(1992).

303. (IL2) Interleukin 2. 31 members

304. (ILVD EDD) Dihydroxy-acid and 6-phosphogluconate dehydratases. Two dehydratases have been shown [1] to be evolutionary related: - Dihydroxy-acid dehydratase (EC 4.2.1.9) (gene ilvD or ILV3) which catalyzes the fourth step in the biosynthesis of isoleucine and valine, the dehydratation of 2,3-dihydroxy-isovaleic acid into alpha-ketoisovaleric acid. - 6-phosphogluconate dehydratase (EC 4.2.1.12) (gene edd) which catalyzes the first step in the Entner-Doudoroff pathway, the dehydratation of 6-phospho- D-gluconate into 6-phospho-2-dehydro-3-deoxy-D-gluconate. - Escherichia coli hypothetical protein yjhG. Both enzymes are proteins of about 600 amino acid residues. Two highly conserved regions have been developed as signature patterns. The first pattern is located in the N-terminal part and contains a cysteine that could be involved in the binding of a 2Fe-2S iron-sulfur cluster [2]. The second pattern is located in the C-terminal half.

Consensus pattern: C-D-K-x(2)-P-[GA]-x(3)-[GA] [The C could be a 2Fe-2S ligand]
Consensus pattern: [SA]-L-[LIVM]-T-D-[GA]-R-[LIVMF]-S-[GA]-[GAV]-[ST]-

[1] Egan S.E., Fliege R., Tong S., Shibata A., Wolf R.E. Jr., Conway T. J. Bacteriol. 174:4638-4646(1992).[2] Velasco J.A., Cansado J., Pena M.C., Kawakami T., Laborda J., Notario V. Gene 137:179-185(1993).

305. IMP dehydrogenase / GMP reductase signature

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1].Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides. IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

Consensus pattern: [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue]-

[1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).

[2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).

[3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).

306. (IPPc) Inositol polyphosphate phosphatase family, catalytic domain

[1] York JD, Ponder JW, Chen ZW, Mathews FS, Majerus PW; Biochemistry 1994;33:13164-13171. [2] Jefferson AB, Auethavekiat V, Pot DA, Williams LT, Majerus PW; J Biol Chem 1997;272:5983-5988. [3] Zhang X, Jefferson AB, Auethavekiat V, Majerus PW; Proc Natl Acad Sci U S A 1995;92:4853-4856. [4] York JD, Majerus PW. Proc Natl Acad Sci U S A 1990;87:9548-9552. [5] Neuwald AF, York JD, Majerus PW; FEBS Lett 1991;294:16-18.

307. IQ calmodulin-binding motif

[1] Xie X, Harrison DH, Schlichting I, Sweet RM, Kalabokis VN, Szent-Gyorgyi AG, Cohen C; Nature 1994;368:306-312.
[2] Rhoads AR, Friedberg F; FASEB J 1997;11:331-340.

308. Inosine-uridine preferring nucleoside hydrolasefamily signature (IU nuc hydro)
Inosine-uridine preferring nucleoside hydrolase (EC 3.2.2.1) (IU-nucleosidehydrolase or IUNH) is an enzyme first identified in protozoan [1] that catalyzes the hydrolysis of all of the commonly occuring purine and pyrimidine nucleosides into ribose and the associated base, but has a preference for inosine and uridine as substrates. This enzyme is important for these parasitic organisms, which are deficient in de novo synthsis of purines, to salvage the host purine nucleosides. IUNH from Crithidia fasciculata has been sequenced and characterized, it is an homotetrameric enzyme of subunits of 34 Kd. An histidine has been shown to be important for the catalytic mechanism, it acts a proton donor to activate the hypoxanthine leaving group. IUNH is evolutionary related to a number of uncharacterized proteins from various biological sources, notably: - Escherichia coli hypothetical protein yaaF. - Escherichia coli hypothetical protein ybeK. - Escherichia coli hypothetical protein yeiK. - Fission yeast hypothetical protein SpAC17G8.02. - Yeast hypothetical protein YDR400w. - An hypothetical protein from the archaebacteria Desulfurolobus ambivalens. As a signature pattern for these proteins, a highly conserved region was selected located in the N-terminal extremity. This region contains four conserved aspartates that have been shown [2] to be located in the active site cavity.

Consensus pattern: D-x-D-[PT]-[GA]-x-D-D-[TAV]-[VI]-A -

[1] Gopaul D.N., Meyer S.L., Degano M., Sacchettini J.C., Schramm V.L. Biochemistry 35:5963-5970(1996).
[2] Degano M., Gopaul D.N., Scapin G., Schramm V.L., Sacchettini J.C. Biochemistry 35:5971-5981(1996).

309. (Insulinase)
Insulinase family, zinc-binding region signature

(aka Peptidase_M16)

A number of proteases dependent on divalent cations for their activity have been shown [1,2] to belong to one family, on the basis of sequence similarity. These enzymes are listed below.

- Insulinase (EC 3.4.24.56) (also known as insulysin or insulin-degrading enzyme or IDE), a cytoplasmic enzyme which seems to be involved in the cellular processing of insulin, glucagon and other small polypeptides.
- Escherichia coli protease III (EC 3.4.24.55) (pitrilysin) (gene ptr), a periplasmic enzyme that degrades small peptides.
- Mitochondrial processing peptidase (EC 3.4.24.64) (MPP). This enzyme removes the transit peptide from the precursor form of proteins imported from the cytoplasm across the mitochondrial inner membrane. It is composed of two nonidentical homologous subunits termed alpha and beta. The beta subunit seems to be catalytically active while the alpha subunit has probably lost its activity.
- Nardilysin (EC 3.4.24.61) (N-arginine dibasic convertase or NRD convertase) this mammalian enzyme cleaves peptide substrates on the N-terminus of Arg residues in dibasic stretches.
- Klebsiella pneumoniae protein pqqF. This protein is required for the biosynthesis of the coenzyme pyrrolo-quinoline-quinone (PQQ). It is thought to be protease that cleaves peptide bonds in a small peptide (gene pqqA) thus providing the glutamate and tyrosine residues necessary for the synthesis of PQQ.
- Yeast protein AXL1, which is involved in axial budding [3].
- Eimeria bovis sporozoite developmental protein.
- Escherichia coli hypothetical protein yddC and HI1368, the corresponding Haemophilus influenzae protein.
- Bacillus subtilis hypothetical protein ymxG.
- Caenorhabditis elegans hypothetical proteins C28F5.4 and F56D2.1.

It should be noted that in addition to the above enzymes, this family also includes the core proteins I and II of the mitochondrial bc1 complex (also called cytochrome c reductase or complex III), but the situation as to the activity or lack of activity of these subunits is quite complex:

- In mammals and yeast, core proteins I and II lack enzymatic activity.
- In Neurospora crassa and in potato core protein I is equivalent to the beta subunit of MPP.
- In Euglena gracilis, core protein I seems to be active, while subunit II is inactive.

These proteins do not share many regions of sequence similarity; the most noticeable is in the N-terminal section. This region includes a conserved histidine followed, two residues later by a glutamate and another histidine. In pitrilysin, it has been shown [4] that this H-x-x-E-H motif is involved in enzyme activity; the two histidines bind zinc and the glutamate is necessary for catalytic activity. Non active members of this family have lost from one to three of these active site residues. We developed a signature pattern that detect active members of this family as well as some inactive members.

Consensus pattern G-x(8,9)-G-x-[STA]-H-[LIVMFY]-[LIVMC]-[DERN]-[HRKL]-[LMFAT]-x-[LFSTH]-x-[GSTAN]-[GST] [The two H are zinc ligands] [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL active members as well as all MPP alpha subunits and core II subunits. Does not detect inactive core I subunits.

Note: these proteins belong to family M16 in the classification of peptidases [5].

[1] Rawlings N.D., Barrett A.J. Biochem. J. 275:389-391(1991).

[2] Braun H.-P., Schmitz U.K. Trends Biochem. Sci. 20:171-175(1995).

[3] Becker A.B., Roth R.A. Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).

[4] Fujita A., Oka C., Arikawa Y., Katagai T., Tonouchi A., Kuhara S., Misumi Y. Nature 372:567-570(1994).

[5] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

310. Involucrin repeat

Eckert RL, Yaffe MB, Crish JF, Murthy S, Rorke EA, Welter JF, J Invest Dermatol 1993;100:613-617.

311. Isochorismatase family. This family are hydrolase enzymes.

Romao MJ, Turk D, Gomis-Ruth FX, Huber R, Schumacher G, Mollering H, Russmann L, J Mol Biol 1992;226:1111-1130.

312. Inositol monophosphatase family signatures (inositol_P)

It has been shown [1] that several proteins share two sequence motifs. Two of these proteins are enzymes of the inositol phosphate second messenger signaling pathway: - Vertebrate and plants inositol monophosphatase (EC 3.1.3.25). - Vertebrate inositol polyphosphate 1-phosphatase (EC 3.1.3.57).The function of the other proteins is not yet clear: - Bacterial protein cysQ. CysQ could help to control the pool of PAPS (3'-phosphoadenoside 5'-phosphosulfate), or be useful in sulfite synthesis. - Escherichia coli protein suhB. Mutations in suhB results in the enhanced synthesis of heat shock sigma factor (htpR). - Neurospora crassa protein Qa-X. Probably involved in quinate metabolism. - Emericella nidulans protein qutG. Probably involved in quinate metabolism. - Yeast protein HAL2/MET22 [2] involved in salt tolerance as well as methionine biosynthesis. - Yeast hypothetical hypothetical protein YHR046c. - Caenorhabditis elegans hypothetical protein F13G3.5. - A Rhizobium leguminosarum hypothetical protein encoded upstream of the pss gene for exopolysaccharide synthesis. - Methanococcus jannaschii hypothetical protein MJ0109.It is suggested [1] that these proteins may act by enhancing the synthesis or degradation of phosphorylated messenger molecules. From the X-ray structure of human inositol monophosphatase [3], it seems that some of the conserved residues are involved in binding a metal ion and/or the phosphate group of the substrate.

Consensus pattern: [FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FY]-x-[HKRNSTY] [The first D and the T bind a metal ion]-
Consensus pattern: [WV]-D-x-[AC]-[GSA]-[GSAPV]-x-[LIVACP]-[LIV]-[LIVAC]-x(3)-[GH]-[GA]-

[1] Neuwald A.F., York J.D., Majerus P.W. FEBS Lett. 294:16-18(1991).
[2] Glaeser H.-U., Thomas D., Gaxiola R., Montrichard F., Surdin-Kerjan Y., Serrano R. EMBO J. 12:3105-3110(1993).

[3] Bone R., Springer J.P., Atack J.R. Proc. Natl. Acad. Sci. U.S.A. 89:10031-10035(1992).

313. Ion transport protein

This family contains Sodium, Potassium, Calcium ion channel This family is 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane. A bacterial structure of the protein is known for the last two helices but is not the Pfam family due to it lacking the first four helices

314. Isocitrate and isopropylmalate dehydrogenases signature (isodh)

Isocitrate dehydrogenase (IDH) [1,2] is an important enzyme of carbohydrate metabolism which catalyzes the oxidative decarboxylation of isocitrate into alpha-ketoglutarate. IDH is either dependent on NAD+ (EC 1.1.1.41) or on NADP+(EC 1.1.1.42). In eukaryotes there are at least three isozymes of IDH: two are located in the mitochondrial matrix (one NAD+-dependent, the other NADP+-dependent), while the third one (also NADP+-dependent) is cytoplasmic. In Escherichia coli the activity of a NADP+-dependent form of the enzyme is controlled by the phosphorylation of a serine residue; the phosphorylated form of IDH is completely inactivated. 3-isopropylmalate dehydrogenase (EC 1.1.1.85) (IMDH) [3,4] catalyzes the third step in the biosynthesis of leucine in bacteria and fungi, the oxidative decarboxylation of 3-isopropylmalate into 2-oxo-4-methylvalerate. Tartrate dehydrogenase (EC 1.1.1.93) [5] catalyzes the reduction of tartrate to oxaloglycolate. These enzymes are evolutionary related [1,3,4,5]. The best conserved region of these enzymes is a glycine-rich stretch of residues located in the C-terminal section. This region was used as a signature pattern.

Consensus pattern: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-[STG]-[LIVMPA]-G-[LIVMF]-

[1] Hurley J.H., Thorsness P.E., Ramalingam V., Helmers N.H., Koshland D.E. Jr., Stroud R.M. Proc. Natl. Acad. Sci. U.S.A. 86:8635-8639(1989).

[2] Cupp J.R., McAlister-Henn L. J. Biol. Chem. 266:22199-22205(1991).

[3] Imada K., Sato M., Tanaka N., Katsube Y., Matsuura Y., Oshima T. J. Mol. Biol. 222:725-738(1991).

[4] Zhang T., Koshland D.E. Jr. Protein Sci. 4:84-92(1995).

[5] Tipton P.A., Beecher B.S. Arch. Biochem. Biophys. 313:15-21(1994).

315. Jacalin-like lectin domain.

Proteins containing this domain are lectins. It is found in 1 to 6 copies in these proteins. The domain is also found in the animal prostatic spermine-binding protein (Swiss:P15501).

[1] Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M; Nat Struct Biol 1996;3:596-603.

316. KH domain

KH motifs probably bind RNA directly. Auto antibodies to Nova, a KH domain protein, cause paraneoplastic opsoclonus ataxia.

[1] Burd CG, Dreyfuss G, Science 1994;265:615-621.

[2] Musco G, Stier G, Joseph C, Castiglione Morelli MA, Nilges M, Gibson TJ, Pastore A, Cell 1996;85:237-245.

317. Kelch motif

The kelch motif was initially discovered in Kelch (Swiss:Q04652). In this protein there are six copies of the motif. It has been shown that Swiss:Q04652 is related to Galactose Oxidase [1] for which a structure has been solved [2]. The kelch motif forms a beta sheet. Several of these sheets associate to form a beta propeller structure as found in neur,

[1] Bork P, Doolittle RF, J Mol Biol 1994;236:1277-1282. [2] Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen, JN, Yadav KD, Knowles PF, Nature 1991;350:87-90.

318. Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature

The soybean trypsin inhibitor (Kunitz) family [1] is one of the numerous families of proteinase inhibitors. It comprise plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases as well as some proteins that are probably involved in seed storage. This family is currently known to group the following proteins: - Trypsin inhibitors A, B, C, KTI1, and KTI2 from soybean. - Trypsin inhibitor DE3 from coral beans (Erythrina sp.). - Trypsin inhibitor DE5 from sandal bead tree. - Trypsin inhibitors 1A (WTI-1A), 1B (WTI-1B), and 2 (WTI-2) from goa bean. - Trypsin inhibitor from Acacia confusa. - Trypsin inhibitor from silk tree. - Chymotrypsin inhibitor 3 (WCI-3) from goa bean. - Cathepsin D inhibitors PDI and NDI from potato [2], which inhibit both cathepsin D (aspartic proteinase) and trypsin. - Alpha-amylase/subtilisin inhibitors from barley and wheat. - Albumin-1 (WBA-1) from goa bean seeds [3]. - Miraculin from Richadella dulcifica [4], a sweet taste protein. - Sporamin from sweet potato [5], the major tuberous root protein. - Thiol proteinase inhibitor PCPI 8.3 (P340) from potato tuber [6]. - Wound responsive protein gwin3 from poplar tree [7]. - 21 Kd seed protein from cocoa [8].All these proteins contain from 170 to 200 amino acid residues and one or twointrachain disulfide bonds. The best conserved region is found in their N-terminal section and is used as a signature pattern

Consensus pattern: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM] -

[1] Laskowski M., Kato I. Annu. Rev. Biochem. 49:593-626(1980).

[2] Ritonja A., Krizaj I., Mesko P., Kopitar M., Lucovnik P., Strukelj B., Pungercar J., Buttle D.J., Barrett A.J., Turk V. FEBS Lett. 267:13-15(1990).

[3] Kortt A.A., Strike P.M., de Jersey J. Eur. J. Biochem. 181:403-408(1989).

[4] Theerasilp S., Hitotsuya H., Nakajo S., Nakaja K., Nakamura Y., Kurihara Y. J. Biol. Chem. 264:6655-6659(1989).

[5] Hattori T., Yoshida N., Nakamura K. Plant Mol. Biol. 13:563-572(1989).

[6] Krizaj I., Drobnic-Kosorok M., Brzin J., Jerala R., Turk V. FEBS Lett. 333:15-20(1993).

[7] Bradshaw H.D., Hollick J.B., Parsons T.J., Clarke H.R.G., Gordon M.P. Plant Mol. Biol. 14:51-59(1989).

[8] Tai H., McHenry L., Fritz P.J., Furtek D.B. Plant Mol. Biol. 16:913-915(1991).

319. Beta-ketoacyl synthases active site

Beta-ketoacyl-ACP synthase (KAS) [1] is the enzyme that catalyzes the condensation of malonyl-ACP with the growing fatty acid chain. It is found as a component of the following enzymatic systems: - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to different enzymatic activities; beta-ketoacyl synthase is one of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the beta-ketoacyl synthase domain is located in the C-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional chain; the beta-ketoacyl synthase domain is located in the N-terminal section [2]. - The multifunctional 6-methysalicylic acid synthase (MSAS) from Penicillium patulum [3]. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which has a KAS domain in its N-terminal section. - Polyketide antibiotic synthase enzyme systems. Polyketides are secondary metabolites produced by microorganisms and plants from simple fatty acids. KAS is one of the components involved in the biosynthesis of the Streptomyces polyketide antibiotics granatacin [4], tetracenomycin C [5] and erythromycin. - Emericella nidulans multifunctional protein Wa. Wa is involved in the biosynthesis of conidial green pigment. Wa is protein of 216 Kd that contains a KAS domain. - Rhizobium nodulation protein nodE, which probably acts as a beta-ketoacyl synthase in the synthesis of the nodulation Nod factor fatty acyl chain. - Yeast mitochondrial protein CEM1. The condensation reaction is a two step process: the acyl component of an activated acyl primer is transferred to a cysteine residue of the enzyme and is then condensed with an activated malonyl donor with the concomitant release of carbon dioxide. The sequence around the active site cysteine is well conserved and can be used as a signature pattern.

Consensus pattern: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF] [C is the active site residue]

[1] Kauppinen S., Siggaard-Andersen M., von Wettstein-Knowles P. Carlsberg Res. Commun. 53:357-370(1988).

[2] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991).

[3] Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E. Eur. J. Biochem. 192:487-498(1990).

[4] Bibb M.J., Biro S., Motamedi H., Collins J.F., Hutchinson C.R. EMBO J. 8:2727-2736(1989).

[5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).

320. Kinesin motor domain signature and profile

Kinesin [1,2,3] is a microtubule-associated force-producing protein that mayplay a role in organelle transport. Kinesin is an oligomeric complex composedof two heavy chains and two light chains. The kinesin motor activity isdirected toward the microtubule's plus end.The heavy chain is composed of three structural domains: a large globular N-terminal domain which is responsible for the motor activity of kinesin (it isknown to hydrolyze ATP, to bind and move on microtubules), a central alpha-helical coiled coil domain that mediates the heavy chain dimerization; and asmall globular C-terminal domain which interacts with other proteins (such asthe kinesin light chains), vesicles and membranous organelles.A number of proteins have been recently found that contain a domain similarto that of the kinesin 'motor' domain [1,4,E1]: - Drosophila claret segregational protein (ncd). Ncd is required for normal chromosomal segregation in meiosis, in females, and in early mitotic divisions of the embryo. The ncd motor activity is directed toward the microtubule's minus end. - Drosophila kinesin-like protein (nod). Nod is required for the distributive chromosome segregation of nonexchange chromosomes during meiosis. - Human CENP-E [4]. CENP-E is a protein that associates with kinetochores during chromosome congression, relocates to the spindle midzone at anaphase, and is quantitatively discarded at the end of the cell division. CENP-E is probably an important motor molecule in chromosome movement and/ or spindle elongation. - Human mitotic kinesin-like protein-1 (MKLP-1), a motor protein whose activity is directed toward the microtubule's plus end. - Yeast KAR3 protein, which is essential for yeast nuclear fusion during mating. KAR3 may mediate microtubule sliding during nuclear fusion and possibly mitosis. - Yeast CIN8 and KIP1 proteins which are required for the assembly of the mitotic spindle. Both proteins seem to interact with spindle microtubules to produce an outwardly directed force acting upon the poles. - Fission yeast cut7 protein, which is essential for spindle body duplication during mitotic division. - Emericella nidulans bimC, which plays an important role in nuclear division. - Emericella nidulans klpA. -

Caenorhabditis elegans unc-104, which may be required for the transport of substances needed for neuronal cell differentiation. - Caenorhabditis elegans osm-3. - Xenopus Eg5, which may be involved in mitosis. - Arabidopsis thaliana KatA, KatB and katC. - Chlamydomonas reinhardtii FLA10/KHP1 and KLP1. Both proteins seem to play a role in the rotation or twisting of the microtubules of the flagella. - Caenorhabditis elegans hypothetical protein T09A5.2.The kinesin motor domain is located in the N-terminal part of most of theabove proteins, with the exception of KAR3, klpA, and ncd where it is locatedin the C-terminal section.The kinesin motor domain contains about 330 amino acids. An ATP-binding motifof type A is found near position 80 to 90, the C-terminal half of the domainis involved in microtubule-binding. The signature pattern for that domain isderived from a conserved decapeptide inside the microtubule-binding part.

Consensus pattern: [GSA]-[KRHPSTQVM]-[LIVMF]-x-[LIVMF]-[IVC]-D-L-[AH]-G-[SAN]-E

[1] Bloom G.S., Endow S.A. Protein Prof. 2:1109-1171(1995).
[2] Vallee R.B., Shpetner H.S. Annu. Rev. Biochem. 59:909-932(1990).
[3] Brady S.T. Trends Cell Biol. 5:159-164(1995).
[4] Endow S.A. Trends Biochem. Sci. 16:221-225(1991).[E1]

321. Ribosomal protein L15 signature

Ribosomal protein L15 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L15 is known to bind the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L15. - Plant chloroplast L15 (nuclear-encoded). - Archaebacterial L15. - Vertebrate L27a. - Tetrahymena thermophila L29. - Fungi L27a (L29, CRP-1, CYH2).L15 is a protein of 144 to 154 amino-acid residues. As a signature pattern, a conserved region was selected in the C-terminal section of these proteins.

Consensus pattern: K-[LIVM](2)-[GASL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x- [LIVMF]-x(3,4)-[LIVMFCA]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G

[1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

322. LBP / BPI / CETP family signature

The following mammalian lipid-binding serum glycoproteins belong to the same family [1,2,3]: - Lipopolysaccharide-binding protein (LBP). LBP binds to the lipid A moiety of bacterial lipopolysaccharides (LPS), a glycolipid present in the outer membrane of all Gram-negative bacteria. The LBP/LPS complex seems to interact with the CD14 receptor and may be responsible for the secretion of alpha-TNF. - Bactericidal permeability-increasing protein (BPI). Like LBP, BPI binds LPS and has a cytotoxic activity on Gram-negative bacteria. - Cholesteryl ester transfer protein (CETP). CETP is involved in the transfer of insoluble cholesteryl esters in reverse cholesterol transport. - Phospholipid transfer protein (PLTP). May play a key role in extracellular phospholipid transport and modulation of HDL particles. These proteins are structurally related and share many regions of sequencesimilarities. As a signature pattern one of these regions was selected, which is located in the N-terminal section of these proteins; a region which could be involved in the binding to the lipids [2].

Consensus pattern: [PA]-[GA]-[LIVMC]-x(2)-R-[IV]-[ST]-x(3)-L-x(5)-[EQ]-x(4)- [LIVM]-[EQK]-x(8)-P

[1] Schumann R.R., Leong S.R., Flaggs G.W., Gray P.W., Wright S.D., Mathison J.C., Tobias P.S., Ulevitch R.J. Science 249:1429-1431(1990).

[2] Gray P.W., Flaggs G., Leong S.R., Gumina R.J., Weiss J., Ooi C.E., Elsbach P. J. Biol. Chem. 264:9505-9509(1989).

[3] Day J.R., Albers J.J., Lofton-Day C.E., Gilbert T.L., Ching A.F.T., Grant F.J., O'Hara P.J., Marcovina S.M., Adolphson J.L. J. Biol. Chem. 269:9388-9391(1994).

323. LIM domain signature and profile

Recently [1,2] a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues. These proteins are: - Caenorhabditis elegans mec-3; a protein required for the differentiation of the set of six touch receptor neurons in this nematode. - Caenorhabditis elegans lin-11; a protein required for the asymmetric division of vulval blast cells. - Vertebrate insulin gene enhancer binding protein isl-1. Isl-1 binds to one of the two cis-acting protein-binding domains of the insulin gene. - Vertebrate homeobox

proteins lim-1, lim-2 (lim-5) and lim3. - Vertebrate lmx-1, which acts as a transcriptional activator by binding to the FLAT element; a beta-cell-specific transcriptional enhancer found in the insulin gene. - Mammalian LH-2, a transcriptional regulatory protein involved in the control of cell differentiation in developing lymphoid and neural cell types. - Drosophila protein apterous, required for the normal development of the wing and halter imaginal discs. - Vertebrate protein kinases LIMK-1 and LIMK-2. - Mammalian rhombotins. Rhombotin 1 (RBTN1 or TTG-1) and rhombotin-2 (RBTN2 or TTG-2) are proteins of about 160 amino acids whose genes are disrupted by chromosomal translocations in T-cell leukemia. - Mammalian and avian cysteine-rich protein (CRP), a 192 amino-acid protein of unknown function. Seems to interact with zyxin. - Mammalian cysteine-rich intestinal protein (CRIP), a small protein which seems to have a role in zinc absorption and may function as an intracellular zinc transport protein. - Vertebrate paxillin, a cytoskeletal focal adhesion protein. - Mouse testin. Mouse testin should not be confused with rat testin which is a thiol protease homolog. - Sunflower pollen specific protein SF3. - Chicken zyxin. Zyxin is a low-abundance adhesion plaque protein which has been shown to interact with CRP. - Yeast protein LRG1 which is involved in sporulation [4]. - Yeast rho-type GTPase activating protein RGA1/DBM1. - Caenorhabditis elegans homeobox protein ceh-14. - Caenorhabditis elegans homeobox protein unc-97. - Yeast hypothetical protein YKR090w. - Caenorhabditis elegans hypothetical proteins C28H8.6.These proteins generally have two tandem copies of a domain, called LIM (forLin-11 Isl-1 Mec-3) in their N-terminal section. Zyxin and paxillin areexceptions in that they contains respectively three and four LIM domains attheir C-terminal extremity. In apterous, isl-1, LH-2, lin-11, lim-1 to lim-3,lmx-1 and ceh-14 and mec-3 there is a homeobox domain some 50 to 95 amino acids after theLIM domains.In the LIM domain, there are seven conserved cysteine residues and ahistidine. The arrangement followed by these conserved residues is C-x(2)-C-x(16,23)-H-x(2)-[CH]-x(2)-C-x(2)-C-x(16,21)-C-x(2,3)-[CHD]. The LIM domainbinds two zinc ions [5]. LIM does not bind DNA, rather it seems to act asinterface for protein-protein interaction. A pattern was developed that spans the first half of the LIM domain.

Consensus pattern: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)- [LIVMF] [The 5 C's and the H bind zinc]

[1] Freyd G., Kim S.K., Horvitz H.R. Nature 344:876-879(1990).

[2] Baltz R., Evrard J.-L., Domon C., Steinmetz A. Plant Cell 4:1465-1466(1992).

[3] Sanchez-Garcia I., Rabbitts T.H. Trends Genet. 10:315-320(1994).

[4] Mueller A., Xu G., Wells R., Hollenberg C.P., Piepersberg W. Nucleic Acids Res. 22:3151-3154(1994).

[5] Michelsen J.W., Schmeichel K.L., Beckerle M.C., Winge D.R. Proc. Natl. Acad. Sci. U.S.A. 90:4404-4408(1993).

324. (LRR) Leucine Rich Repeat

CAUTION: This Pfam may not find all Leucine Rich Repeats in a protein. Leucine Rich Repeats are short sequence motifs present in a number of proteins with diverse functions and cellular locations. These repeats are usually involved in protein-protein interactions. Each Leucine Rich Repeat is composed of a beta-alpha unit. These units form elongated non-globular structures. Leucine Rich Repeats are often flanked by cysteine rich domains.

Number of members: 3017

[1] The leucine-rich repeat: a versatile binding motif. Kobe B, Deisenhofer J; Trends Biochem Sci 1994;19:415-421. [2] Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Kobe B, Deisenhofer J; Nature 1993;366:751-756.

325. Plant lipid transfer protein family signature (LTP)

Plant cells contain proteins, called lipid transfer proteins (LTP) [1,2,3], which are able to facilitate the transfer of phospholipids and other lipidsacross membranes. These proteins, whose subcellular location is not yet known, could play a major role in membrane biogenesis by conveying phospholipids such as waxes or cutin from their site of biosynthesis to membranes unable to form these lipids. Plant LTP's are proteins of about 9 Kd (90 amino acids) which contain eight conserved cysteine residues all involved in disulfide bridges, as shown in the following schematic representation.

+----------------------+ | +------+ | | | | | ***************

xCxxxxCxxxxxxCCxxxxxxxxCxCxxxxxxxxxxxCxxxxxxCxx | | | | +--------|------------+ | +---
-----------------+

'C': conserved cysteine involved in a disulfide bond.

'*': position of the pattern.

Consensus pattern: [LIVM]-[PA]-x(2)-C-x-[LIVM]-x-[LIVM]-x-[LIVMFY]-x-[LIVM]-[ST]-x(3)-[DN]-C-x(2)-[LIVM] [The two C's are involved in disulfide bonds]

[1] Wirtz K.W.A. Annu. Rev. Biochem. 60:73-99(1991).
[2] Arondel V., Kader J.C. Experientia 46:579-585(1990).
[3] Ohlrogge J.B., Browse J., Somerville C.R. Biochim. Biophys. Acta 1082:1-26(1991).

326. (LAMP) Lysosome-associated membrane glycoproteins signatures

Lysosome-associated membrane glycoproteins (lamp) [1] are integral membrane proteins, specific to lysosomes, and whose exact biological function is not yet clear. Structurally, the lamp proteins consist of two internally homologous lysosome-luminal domains separated by a proline-rich hinge region; at the C-terminal extremity there is a transmembrane region followed by a very short cytoplasmic tail. In each of the duplicated domains, there are two conserved disulfide bonds. This structure is schematically represented in the figure below. +-----+ +-----+ +-----+ +-----+ | | | | | | | | xCxxxxxCxxxxxxxxxxxxCxxxxxCxxxxxxxxxxCxxxxxCxxxxxxxxxxxxCxxxxxCxxxxxxxxx <------------------------><Hinge><------------------------><TM><C> In mammals, there are two closely related types of lamp: lamp-1 and lamp-2. In chicken lamp-1 is known as LEP100.The macrophage protein CD68 (or macrosialin) [2] is a heavily glycosylatedintegral membrane protein whose structure consists of a mucin-like domain followed by a proline-rich hinge; a single lamp-like domain; a transmembrane region and a short cytoplasmic tail. Two signature patterns for this family of proteins were developed. The first oneis centered on the first conserved cysteine of the duplicated domains. The second corresponds to a region that includes the extremity of the second domain, the totality of the transmembrane region and the cytoplasmic tail.

Consensus pattern: [STA]-C-[LIVM]-[LIVMFYW]-A-x-[LIVMFYW]-x(3)-[LIVMFYW]-x(3)-Y [C is involved in a disulfide bond] –
Consensus pattern: C-x(2)-D-x(3,4)-[LIVM](2)-P-[LIVM]-x-[LIVM]-G-x(2)-[LIVM]- x-G-[LIVM](2)-x-[LIVM](4)-A-[FY]-x-[LIVM]-x(2)-[KR]-[RH]- x(1,2)-[STAG](2)-Y-[EQ] [C is involved in a disulfide bond]

[1] Fukuda M. J. Biol. Chem. 266:21327-21330(1991).

[2] Holness C.L., da Silva R.P., Fawcett J., Gordon S., Simmons D.L. J. Biol. Chem. 268:9661-9666(1993).

327. Lipolytic enzymes "G-D-S-L" family, serine active site

Recently [1], a family of lipolytic enzymes has been characterized. This family currently consist of the following proteins:

- Aeromonas hydrophila lipase/phosphatidylcholine-sterol acyltransferase.
- Xenorhabdus luminescens lipase 1.
- Vibrio mimicus arylesterase.
- Escherichia coli acyl-coA thioesterase I (gene tesA).
- Vibrio parahaemolyticus thermolabile hemolysin/atypical phospholipase.
- Rabbit phospholipase AdRab-B, an intestinal brush border protein with esterase and phospholipase A/lysophospholipase activity that could be involved in the uptake of dietary lipids. AdRab-B contains four repeats of about 320 amino acids.
- Arabidopsis thaliana and Brassic napus anther-specific proline-rich protein APG.
- A Pseudomonas putida hypothetical protein in trpE-trpG intergenic region. A serine has been identified a part of the active site in the Aeromonas, Vibrio mimicus and Escherichia coli enzymes. It is located in a conserved sequence motif that can be used as a signature pattern for these proteins.

-Consensus pattern: [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G
[S is the active site residue]

328. (Lipoprotein 4) Prokaryotic membrane lipoprotein lipid attachment site
In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signalpeptidase II). The peptidase recognizes a conserved sequence and cuts upstreamof a cysteine residue to which a glyceride-fatty acid lipid is attached [1].Some of the proteins known to undergo such processing currently include (forrecent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). -

Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS. - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaon is [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion).From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification was derived.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).

[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).

[3] von Heijne G. Protein Eng. 2:531-534(1989).

[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

329. (Lopoprotein 5) Prokaryotic membrane lipoprotein lipid attachment site. In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1].Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). - Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS. - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlp ABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion).From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification have been developed.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and

35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).[3] von Heijne G. Protein Eng. 2:531-534(1989).[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

330. (Lum binding) Riboflavin synthase alpha chain family Lum-binding site signature The following proteins have been shown [1,2] to be structurally and evolutionary related: - Riboflavin synthase alpha chain (RS-alpha) (gene ribC in Escherichia coli, ribB in Bacillus subtilis and Photobacterium leiognathi, RIB5 in yeast). This enzyme synthesizes riboflavin from two moles of 6,7- dimethyl-8-(1'-D-ribityl)lumazine (Lum), a pteridine-derivative. - Photobacterium phosphoreum lumazine protein (LumP) (gene luxL). LumP is a protein that modulates the color of the bioluminescence emission of bacterial luciferase. In the presence of LumP, light emission is shifted to higher energy values (shorter wavelength). LumP binds non-covalently to 6,7-dimethyl-8-(1'-D-ribityl) lumazine. - Vibrio fischeri yellow fluorescent protein (YFP) (gene luxY). Like LumP, YFP modulates light emission but towards a longer wavelength. YFP binds non-covalently to FMN. These proteins seem to have evolved from the duplication of a domain of about100 residues. In its C-terminal section, this domain contains a conserved motif [KR]-V-N-[LI]-E which has been proposed to be the binding site for Lum.RS-alpha which binds two molecules of Lum has two perfect copies of this motif, while LumP which binds one molecule of Lum, has a Glu instead of Lys/Arg in the first position of the second copy of the motif. Similarily, YFP, which binds to one molecule of FMN, also seems to have a potentially dysfunctional binding site by substitution of Gly for Glu in the last positionof the first copy of the motif. Our signature pattern includes the Lum-binding motif.

Consensus pattern: [LIVMF]-x(5)-G-[STADNQ]-[KREQIYW]-V-N-[LIVM]-E

[1] O'Kane D.J., Woodward B., Lee J., Prasher D.C. Proc. Natl. Acad. Sci. U.S.A. 88:1100-1104(1991).

[2] O'Kane D.J., Prasher D.C. Mol. Microbiol. 6:443-449(1992).

331. Lysyl oxidase putative copper-binding region signature

Lysyl oxidase (LOX) [1] is an extracellular copper-dependent enzyme that catalyzes the oxidative deamination of peptidyl lysine residues in precursors of various collagens and elastins. The deaminated lysines are then able to form aldehyde cross-links.LOX binds a single copper atom which seems to reside within an octahedral coordination complex which includes at least three histidine ligands. Fourhistidine residues are clustered in a central region of the enzyme. This region is thought to be involved in cooper-binding and is called the 'copper-talon' [1]. This region was used as a signature pattern.

Consensus pattern: W-E-W-H-S-C-H-Q-H-Y-H

[1] Krebs C.J., Krawetz S.A. Biochim. Biophys. Acta 1202:7-12(1993).

332. Metallo-beta-lactamase superfamily (lactamase_B)

[1] : Neuwald AF, Liu JS, Lipman DJ, Lawrence CE, Nucleic Acids Res 1997;25:1665-1677. [2] Carfi A, Pares S, Duee E, Galleni M, Duez C, Frere JM, Dideberg O, EMBO J 1995;14:4914-4921.

333. L-lactate dehydrogenase active site (ldh1)

L-lactate dehydrogenase (EC 1.1.1.27) (LDH) [1] catalyzes the reversible NAD-dependent interconversion of pyruvate to L-lactate. In vertebrate muscles and in lactic acid bacteria it represents the final step in anaerobic glycolysis. This tetrameric enzyme is present in prokaryotic and eukaryotic organisms. Invertebrates there are three isozymes of LDH: the M form (LDH-A), found predominantly in muscle tissues; the H form (LDH-B), found in heart muscle and the X form (LDH-C), found only in the spermatozoa of mammals and birds. In birds and crocodilian eye lenses, LDH-B serves as a structural protein and is known as epsilon-crystallin [2].L-2-hydroxyisocaproate dehydrogenase (EC 1.1.1.-) (L-hicDH) [3] catalyzes the reversible and stereospecific interconversion between 2-ketocarboxylic acids and L-2-hydroxy-carboxylic acids. L-hicDH is evolutionary related to LDH's. As a signature

for LDH's a region was selected that includes a conserved histidine which is essential to the catalytic mechanism.

Consensus pattern: [LIVMA]-G-[EQ]-H-G-[DN]-[ST] [H is the active site residue] -

[1] Abad-Zapatero C., Griffith J.P., Sussman J.L., Rossmann M.G. J. Mol. Biol. 198:445-467(1987).
[2] Hendriks W., Mulders J.W.M., Bibby M.A., Slingsby C., Bloemendal H., de Jong W.W. Proc. Natl. Acad. Sci. U.S.A. 85:7114-7118(1988).
[3] Lerch H.-P., Frank R., Collins J. Gene 83:263-270(1989).

Malate dehydrogenase active site signature (ldh2)
Malate dehydrogenase (EC 1.1.1.37) (MDH) [1,2] catalyzes the interconversion of malate to oxaloacetate utilizing the NAD/NADH cofactor system. The enzyme participates in the citric acid cycle and exists in all aerobic organisms. While prokaryotic organisms contains a single form of MDH, in eukaryotic cells there are two isozymes: one which is located in the mitochondrial matrix and the other in the cytoplasm. Fungi and plants also harbor a glyoxysomal form which functions in the glyoxylate pathway. In plants chloroplast there is an additional NADP-dependent form of MDH (EC 1.1.1.82) which is essential for both the universal C3 photosynthesis (Calvin) cycle and the more specializedC4 cycle. As a signature pattern for this enzyme a region was chosen that includes two residues involved in the catalytic mechanism [3]: an aspartic acid which is involved in a proton relay mechanism, and an arginine which binds the substrate.

Consensus pattern: [LIVM]-T-[TRKMN]-L-D-x(2)-R-[STA]-x(3)-[LIVMFY] [D and R are the active site residues]-

[1] McAlister-Henn L. Trends Biochem. Sci. 13:178-181(1988).
[2] Gietl C. Biochim. Biophys. Acta 1100:217-234(1992).
[3] Birktoft J.J., Rhodes G., Banaszak L.J. Biochemistry 28:6065-6081(1989).
[4] Cendrin F., Chroboczek J., Zaccai G., Eisenberg H., Mevarech M. Biochemistry 32:4308-4313(1993).

334. Legume lectins signatures

Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals). Legume lectins are synthesized as precursor proteins of about 230 to 260 amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal).The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain. Two signature patterns specific to legume lectins have been developed: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain.

Consensus pattern: [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium]-

Consensus pattern: [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST]-

[1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).

[2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).

335. CoA-ligases (ligases- CoA)

This family includes the CoA ligases Succinyl-CoA synthetase alpha: and beta chains, malate CoA ligase and ATP-citrate lyase. Some members of the family utilise ATP others use GTP.

[1] Wolodko WT, Fraser ME, James MN, Bridger WA, J Biol Chem 1994;269:10883-10890.

336. linker histone H1 and H5 family

Linker histone H1 is an essential component of chromatin structure. H1 links nucleosomes into higher order structures Histone H1 is replaced by histone H5 in some cell types.

[1] Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM, Nature 1993;362:219-223.

337. Lipocalin signature (lip1)

Proteins which transport small hydrophobic molecules such as steroids, bilins, retinoids, and lipids share limited regions of sequence homology and a common tertiary structure architecture [1 to 5]. This is an eight stranded antiparallel beta-barrel with a repeated + 1 topology enclosing a internal ligand binding site [1,3]. The name 'lipocalin' has been proposed [5] for this protein family. Proteins known to belong to this family are listed below (references are only provided for recently determined sequences). - Alpha-1-microglobulin (protein HC), which seems to bind porphyrin. - Alpha-1-acid glycoprotein (orosomucoid), which can bind a remarkable array of natural and synthetic compounds [6]. - Aphrodisin which, in hamsters, functions as an aphrodisiac pheromone. - Apolipoprotein D, which probably binds heme-related compounds. - Beta-lactoglobulin, a milk protein whose physiological function appears to bind retinol. - Complement component C8 gamma chain, which seems to bind retinol [7]. - Crustacyanin [8], a protein from lobster carapace, which binds astaxanthin, a carotenoid. - Epididymal-retinoic acid binding protein (E-RABP) [9] involved in sperm maturation. - Insectacyanin, a moth bilin-binding protein, and a related butterfly bilin- binding protein (BBP). - Late Lactation protein (LALP), a milk protein from tammar wallaby [10]. - Neutrophil gelatinase-associated lipocalin (NGAL) (p25) (SV-40 induced 24p3 protein) [11]. - Odorant-binding protein (OBP), which binds odorants. - Plasma retinol-binding proteins (PRBP). - Human pregnancy-associated endometrial alpha-2 globulin. - Probasin (PB), a rat prostatic protein. - Prostaglandin D synthase (EC 5.3.99.2) (GSH-independent PGD synthetase), a lipocalin with enzymatic activity [12]. - Purpurin, a retinal protein which binds retinol and heparin. - Quiescence specific protein p20K from chicken (embryo CH21 protein). - Rodent urinary proteins (alpha-2-microglobulin), which may bind pheromones. - VNSP 1 and 2, putative pheromone transport proteins from mouse vomeronasal organ [13]. - Von Ebner's gland protein (VEGP) [14] (also called tear lipocalin), a mammalian protein which may be involved in taste recognition. - A frog olfactory protein, which may transport odorants. - A protein found in the cerebrospinal fluid of the toad Bufo

Marinus with a supposed function similar to transthyretin in transport across the blood brain barrier [15]. - Lizard's epididymal secretory protein IV (LESP IV), which could transport small hydrophobic molecules into the epididymal fluid during sperm maturation [16]. - Prokaryotic outer-membrane protein blc [17].The sequences of most members of the family, the core or kernal lipocalins, are characterized by three short conserved stretches of residues [3,18].Others, the outlier lipocalin group, share only one or two of these [3,18]. A signature pattern was built around the first, common to all outlier and kernallipocalins, which occurs near the start of the first beta-strand.

Consensus pattern: [DENG]-x-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}- W-[FYWLRH]-x-[LIVMTA]-

Note: it is suggested, on the basis of similarities of structure, function, and sequence, that this family forms an overall superfamily, called the calycins, with the avidin/streptavidin <PDOC00499> and the cytosolic fatty- acid binding proteins <PDOC00188> families [3,19]

[1] Cowan S.W., Newcomer M.E., Jones T.A. Proteins 8:44-61(1990).

[2] Igaraishi M., Nagata A., Toh H., Urade H., Hayaishi N. Proc. Natl. Acad. Sci. U.S.A. 89:5376-5380(1992).

[3] Flower D.R., North A.C.T., Attwood T.K. Protein Sci. 2:753-761(1993).

[4] Godovac-Zimmermann J. Trends Biochem. Sci. 13:64-66(1988).

[5] Pervaiz S., Brew K. FASEB J. 1:209-214(1987).

[6] Kremer J.M.H., Wilting J., Janssen L.H.M. Pharmacol. Rev. 40:1-47(1989).

[7] Haefliger J.-A., Peitsch M.C., Jenne D., Tschopp J. Mol. Immunol. 28:123-131(1991).

[8] Keen J.N., Caceres I., Eliopoulos E.E., Zagalsky P.F., Findlay J.B.C. Eur. J. Biochem. 197:407-417(1991).

[9] Newcomer M.E. Structure 1:7-18(1993).

[10] Collet C., Joseph R. Biochim. Biophys. Acta 1167:219-222(1993).

[11] Kjeldsen L., Johnsen A.H., Sengelov H., Borregaard N. J. Biol. Chem. 268:10425-10432(1993).

[12] Peitsch M.C., Boguski M.S. Trends Biochem. Sci. 16:363-363(1991).

[13] Miyawaki A., Matsushita Y.R., Ryo Y., Mikoshiba T. EMBO J. 13:5835-5842(1994).

[14] Kock K., Ahlers C., Schmale H. Eur. J. Biochem. 221:905-916(1994).

[15] Achen M.G., Harms P.J., Thomas T., Richardson S.J., Wettenhall R.E.H., Schreiber G. J. Biol. Chem. 267:23170-23174(1992).

[16] Morel L., Dufarre J.-P., Depeiges A. J. Biol. Chem. 268:10274-10281(1993).

[17] Bishop R.E., Penfold S.S., Frost L.S., Holtje J.V., Weiner J.H. J. Biol. Chem. 270:23097-23103(1995).

[18] Flower D.R., North A.C.T., Attwood T.K. Biochem. Biophys. Res. Commun. 180:69-74(1991).

[19] Flower D.R. FEBS Lett. 333:99-102(1993).

Cytosolic fatty-acid binding proteins signature (lip2)

A number of low molecular weight proteins which bind fatty acids and other organic anions are present in the cytosol [1,2]. Most of them are structurally related and have probably diverged from a common ancestor. This structure is a ten stranded antiparallel beta-barrel, albeit with a wide discontinuity between the fourth and fifth strands, with a repeated + 1 topology enclosing an internal ligand binding site [2,7]. Proteins known to belong to this family include: - Six, tissue-specific, types of fatty acid binding proteins (FABPs) found in liver, intestine, heart, epidermal, adipocyte, brain/retina. Heart FABP is also known as mammary-derived growth inhibitor (MDGI), a protein that reversibly inhibits proliferation of mammary carcinoma cells. Epidermal FABP is also known as psoriasis-associated FABP [3]. - Insect muscle fatty acid-binding proteins. - Testis lipid binding protein (TLBP). - Cellular retinol-binding proteins I and II (CRBP). - Cellular retinoic acid-binding protein (CRABP). - Gastrotropin, an ileal protein which stimulates gastric acid and pepsinogen secretion. It seems that gastrotropin binds to bile salts and bilirubins. - Fatty acid binding proteins MFB1 and MFB2 from the midgut of the insect Manduca sexta [4].In addition to the above cytosolic proteins, this family also includes: - Myelin P2 protein, which may be a lipid transport protein in Schwann cells. P2 is associated with the lipid bilayer of myelin. - Schistosoma mansoni protein Sm14 [5] which seems to be involved in the transport of fatty acids. - Ascaris suum p18 a secreted protein that may play a role in sequestering potentially toxic fatty acids and their peroxidation products or that may be involved in the maintenance of the impermeable lipid layer of the eggshell. - Hypothetical fatty acid-binding proteins F40F4.2, F40F4.3, F40F4.4 and ZK742.5 from Caenorhabditis elegans. As a signature pattern for these proteins a segment from the N-terminal extremity was use.

Consensus pattern: [GSAIVK]-x-[FYW]-x-[LIVMF]-x(4)-[NHG]-[FY]-[DE]-x- [LIVMFY]-[LIVM]-x(2)-[LIVMAKR]-

Note: it is suggested, on the basis of similarities of structure, function, and sequence, that this family forms an overall superfamily, called the calycins, with the lipocalin <PDOC00187> and avidin/streptavidin <PDOC00499> families [6,7].

[1] Bernier I., Jolles P. Biochimie 69:1127-1152(1987).

[2] Veerkamp J.H., Peeters R.A., Maatman R.G.H.J. Biochim. Biophys. Acta 1081:1-24(1991).

[3] Siegenthaler G., Hotz R., Chatellard-Gruaz D., Didierjean L., Hellman U., Saurat J.-H. Biochem. J. 302:363-371(1994).

[4] Smith A.F., Tsuchida K., Hanneman E., Suzuki T.C., Wells M.A. J. Biol. Chem. 267:380-384(1992).

[5] Moser D., Tendler M., Griffiths G., Klinkert M.-Q. J. Biol. Chem. 266:8447-8454(1991).

[6] Flower D.R., North A.C.T, Attwood T.K. Protein Sci. 2:753-761(1993).

[7] Flower D.R. FEBS Lett. 333:99-102(1993).

338. Lipoxygenases iron-binding region signatures

Lipoxygenases (EC 1.13.11.-) are a class of iron-containing dioxygenases which catalyzes the hydroperoxidation of lipids, containing a cis,cis-1,4-pentadiene structure. They are common in plants where they may be involved in a number of diverse aspects of plant physiology including growth and development, pest resistance, and senescence or responses to wounding [1]. In mammals a number of lipoxygenases isozymes are involved in the metabolism of prostaglandins and leukotrienes [2]. Sequence data is available for the following lipoxygenases: - Plant lipoxygenases (EC 1.13.11.12). Plants express a variety of cytosolic isozymes as well as what seems [3] to be a chloroplast isozyme. - Mammalian arachidonate 5-lipoxygenase (EC 1.13.11.34). - Mammalian arachidonate 12-lipoxygenase (EC 1.13.11.31). - Mammalian erythroid cell-specific 15-lipoxygenase (EC 1.13.11.33).The iron atom in lipoxygenases is bound by four ligands, three of which are histidine residues [4]. Six histidines are conserved in all lipoxygenase sequences, five of them are found clustered in a stretch of 40 amino acids. This region contains two of the three zinc-ligands; the other histidines have been shown [5] to be important for the activity of lipoxygenases. As signatures for this family of enzymes two patterns in the region of the histidine cluster were selected. The first pattern contains the first three conserved histidines and the second pattern includes the fourth and the fifth.

Consensus pattern: H-[EQ]-x(3)-H-x-[LM]-[NQRC]-[GST]-H-[LIVMSTAC](3)-E [The second and third H's bind iron]-
Consensus pattern: [LIVMA]-H-P-[LIVM]-x-[KRQ]-[LIVMF](2)-x-[AP]-H-

[1] Vick B.A., Zimmerman D.C. (In) Biochemistry of plants: A comprehensive treatise, Stumpf P.K., Ed., Vol. 9, pp.53-90, Academic Press, New-York, (1987).
[2] Needleman P., Turk J., Jakschik B.A., Morrison A.R., Lefkowith J.B. Annu. Rev. Biochem. 55:69-102(1986).
[3] Peng Y.L., Shirano Y., Ohta H., Hibino T., Tanaka K., Shibata D. J. Biol. Chem. 269:3755-3761(1994).
[4] Boyington J.C., Gaffney B.J., Amzel L.M. Science 260:1482-1486(1993).
[5] Steczko J., Donoho G.P., Clemens J.C., Dixon J.E., Axelrod B. Biochemistry 31:4053-4057(1992).

339. Fumarate lyases signature (lyase_1)
A number of enzymes, belonging to the lyase class, for which fumarate is a substrate have been shown [1,2] to share a short conserved sequence around a methionine which is probably involved in the catalytic activity of this type of enzymes. These enzymes are: - Fumarase (EC 4.2.1.2) (fumarate hydratase), which catalyzes the reversible hydration of fumarate to L-malate. There seem to be 2 classes of fumarases: class I are thermolabile dimeric enzymes (as for example: Escherichia coli fumC); class II enzymes are thermostable and tetrameric and are found in prokaryotes (as for example: Escherichia coli fumA and fumB) as well as in eukaryotes. The sequence of the two classes of fumarases are not closely related. - Aspartate ammonia-lyase (EC 4.3.1.1) (aspartase), which catalyzes the reversible conversion of aspartate to fumarate and ammonia. This reaction is analogous to that catalyzed by fumarase, except that ammonia rather than water is involved in the trans-elimination reaction. - Arginosuccinase (EC 4.3.2.1) (argininosuccinate lyase), which catalyzes the formation of arginine and fumarate from argininosuccinate, the last step in the biosynthesis of arginine. - Adenylosuccinase (EC 4.3.2.2) (adenylosuccinate lyase) [3], which catalyzes the eight step in the de novo biosynthesis of purines, the formation of 5'-phosphoribosyl-5-amino-4-imidazolecarboxamide and fumarate from 1-(5- phosphoribosyl)-4-(N-succino-carboxamide). That enzyme can also catalyzes the formation of fumarate and AMP from adenylosuccinate. -

Pseudomonas putida 3-carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2) (3-carboxymuconate lactonizing enzyme) (gene pcaB) [4], an enzyme involved in aromatic acids catabolism

Consensus pattern: G-S-x(2)-M-x(2)-K-x-N-

[1] Woods S.A., Shwartzbach S.D., Guest J.R. Biochim. Biophys. Acta 954:14-26(1988).
[2] Woods S.A., Miles J.S., Guest J.R. FEMS Microbiol. Lett. 51:181-186(1988).
[3] Zalkin H., Dixon J.E. Prog. Nucleic Acid Res. Mol. Biol. 42:259-287(1992).
[4] Williams S.E., Woolridge E.M., Ransom S.C., Landro J.A., Babbitt P.C., Kozarich J.W. Biochemistry 31:9768-9776(1992).

340. MCM family signature and profile

Proteins shown to be required for the initiation of eukaryotic DNA replication share a highly conserved domain of about 210 amino-acid residues [1,2,3]. The latter shows some similarities [4] with that of various other families of DNA-dependent ATPases. Eukaryotes seem to possess a family of six proteins that contain this domain. They were first identified in yeast where most of them have a direct role in the initiation of chromosomal DNA replication by interacting directly with autonomously replicating sequences (ARS). They were thus called 'minichromosome maintenance proteins' with gene symbols prefixed by MCM. These six proteins are: - MCM2, also known as cdc19 (in S.pombe) [E1]. - MCM3, also known as DNA polymerase alpha holoenzyme-associated protein P1, RLF beta subunit or ROA. - MCM4, also known as CDC54, cdc21 (in S.pombe) or dpa (in Drosophila). - MCM5, also known as CDC46 or nda4 (in S.pombe). - MCM6, also known as mis5 (in S.pombe). - MCM7, also known as CDC47 or Prolifera (in A.thaliana).This family is also present in archebacteria. In Methanococcus jannaschiithere are four members: MJ0363, MJ0961, MJ1489 and MJECL13.The presence of a putative ATP-binding domain implies that these proteins maybe involved in an ATP-consuming step in the initiation of DNA replication in eukaryotes. As a signature pattern, a perfectly conserved region was selected that represents a special version of the B motif found in ATP-binding proteins.

Consensus pattern: G-[IVT]-[LVAC](2)-[IVT]-D-[DE]-[FL]-[DNST]

[1] Coxon A., Maundrell K., Kearsey S.E. Nucleic Acids Res. 20:5571-5577(1992).

[2] Hu B., Burkhart R., Schulte D., Musahl C., Knippers R. Nucleic Acids Res. 21:5289-5293(1993).

[3] Tye B.-K. Trends Cell Biol. 4:160-166(1994).

[4] Koonin E.V. Nucleic Acids Res. 21:2541-2547(1993).

341. Macrophage migration inhibitory factor family signature (MIF)

A protein called macrophage migration inhibitory factor (MIF) [1] seems to exert an important role in host inflammatory responses. It play a pivotal role in the host response to endotoxic shock and appears to serve as a pituitary "stress" hormone that regulates systemic inflammatory responses. MIF is a secreted protein of 115 residues which is not processed from a larger precursor. D-dopachrome tautomerase [2] is a mammalian cytoplasmic enzyme involved in melanin biosynthesis and that tautomerizes D-dopachrome with concomitant decarboxylation to give 5,6-dihydroxyindole (DHI). It is a protein of 117 residues highly related to MIF. It must be noted that MIF binds glutathione and has been said to be related to glutathione S-transferases. This assertion has been later disproved [3].As a signature pattern for these proteins, a conserved region was selected located in the central section.

Consensus pattern: [DE]-P-C-A-x(3)-[LIVM]-x-S-I-G-x-[LIVM]-G-

[1] Bucala R. Immunol. Lett. 43:23-26(1994).

[2] Odh G., Hindemith A., Rosengren A.-M., Rosengren E., Rorsman H. Biochem. Biophys. Res. Commun. 197:619-624(1993).

[3] Pearson W.R. Protein Sci. 3:525-527(1994).

342. MIP family signature

Recently the sequence of a number of different proteins, that all seem to be transmembrane channel proteins, has been found to be highly related [1 to 4].These proteins are listed below.
- Mammalian major intrinsic protein (MIP). MIP is the major component of lens fiber gap junctions. Gap junctions mediate direct exchange of ions and small molecule from one cell to another. - Mammalian aquaporins [5]. These proteins form water-specific channels that provide the plasma membranes of red cells and kidney proximal and collecting tubules with

high permeability to water, thereby permitting water to move in the direction of an osmotic gradient. - Soybean nodulin-26, a major component of the peribacteroid membrane induced during nodulation in legume roots after Rhizobium infection. - Plants tonoplast intrinsic proteins (TIP). There are various isoforms of TIP: alpha (seed), gamma, Rt (root), and Wsi (water-stress induced). These proteins may allow the diffusion of water, amino acids and/or peptides from the tonoplast interior to the cytoplasm. - Bacterial glycerol facilitator protein (gene glpF), which facilitates the movement of glycerol across the cytoplasmic membrane. - Salmonella typhimurium propanediol diffusion facilitator (gene pduF). - Yeast FPS1, a glycerol uptake/efflux facilitator protein. - Drosophila neurogenic protein 'big brain' (bib). This protein may mediate intercellular communication; it may functions by allowing the transport of certain molecules(s) and thereby sending a signal for an exodermal cell to become an epidermoblast instead of a neuroblast. - Yeast hypothetical protein YFL054c. - A hypothetical protein from the pepX region of lactococcus lactis. The MIP family proteins seem to contain six transmembrane segments. Computer analysis shows that these protein probably arose by a tandem, intragenic duplication event from an ancestral protein that contained three transmembrane segments. As a signature pattern a well conserved region was selected which is located in a probable cytoplasmic loop between the second and third transmembrane regions.

Consensus pattern: [HNQA]-x-N-P-[STA]-[LIVMF]-[ST]-[LIVMF]-[GSTAFY]-

[1] Reizer J., Reizer A., Saier M.H. Jr. CRC Crit. Rev. Biochem. 28:235-257(1993).
[2] Baker M.E., Saier M.H. Jr. Cell 60:185-186(1990).
[3] Pao G.M., Wu L.-F., Johnson K.D., Hoefte H., Chrispeels M.J., Sweet G., Sandal N.N., Saier M.H. Jr. Mol. Microbiol. 5:33-37(1991).
[4] Wistow G.J., Pisano M.M., Chepelinsky A.B. Trends Biochem. Sci. 16:170-171(1991).
[5] Chrispeels M.J., Agre P. Trends Biochem. Sci. 19:421-425(1994).

343. Mandelate racemase / muconate lactonizing enzyme family signatures

Mandelate racemase (EC 5.1.2.2) (MR) and muconate lactonizing enzyme(EC 5.5.1.1) (MLE) are two bacterial enzymes involved in aromatic acid catabolism. They catalyze mechanistically distinct reactions yet they are related at the level of their primary, quaternary (homooctamer) and tertiary structures [1,2].A number of other proteins also seem to be

evolutionary related to these two enzymes. These are: - The various plasmid-encoded chloromuconate cycloisomerases (EC 5.5.1.7). - Escherichia coli protein rspA [3], rspA seems to be involved in the degradation of homoserine lactone (HSL) or of one of its metabolite. - Escherichia coli hypothetical protein ycjG. - Escherichia coli hypothetical protein yidU. - A hypothetical protein from Streptomyces ambofaciens [4]. Two signature patterns have been developed for these enzymes; both contain conserved acidic residues. The second pattern contains an aspartate and a glutamate which are ligands for either a magnesium ion (in MR) or a manganese ion (inMLE).

Consensus pattern: A-x-[SAGCN]-[SAG]-[LIVM]-[DEQ]-x-A-[LA]-x-[DE]-[LIA]-x- [GA]-[KRQ]-x(4)-[PSA]-[LIV]-x(2)-L-[LIVMF]-G-
Consensus pattern: [LIVF]-x(2)-D-x-[NH]-x(7)-[ACL]-x(6)-[LIVMF]-x(7)-[LIVM]- E-[DENQ]-P [D and E bind a divalent metal ion]-

[1] Neidhart D.J., Kenyon G.L., Gerlt J.A., Petsko G.A. Nature 347:692-694(1990).
[2] Petsko G.A., Kenyon G.L., Gerlt J.A., Ringe D., Kozarich J.W. Trends Biochem. Sci. 18:372-376(1993).
[3] Huisman G.W., Kolter R. Science 265:537-539(1994).
[4] Schneider D., Aigle B., Leblond P., Simonet J.M., Decaris B. J. Gen. Microbiol. 139:2559-2567(1993).

344. Merozoite Surface Antigen 2 (MSA-2) family

Thomas AW, Carr DA, Carter JM, Lyon JA, Mol Biochem Parasitol 1990;43:211-220.

345. MSP (Major sperm protein) domain.

Major sperm proteins are involved in sperm motility. These proteins oligomerise to form filaments. Partial matches to this domain are also found in other non MSP proteins. These include Swiss:P40075 and Swiss:P34593.

[1] Bullock TL, Roberts TM, Stewart M, J Mol Biol 1996;263:284-296. [2] King KL, Stewart M, Roberts TM, Seavy M, J Cell Sci 1992;101:847-857.

346. (Matrix) Viral matrix protein. Found in Morbillivirus and paramyxovirus, pneumovirus. Number of members: 105

347. O-methyltransferase (methyltransf)

This family includes a range of O-methyltransferases. These enzymes utilise S-adenosyl methionine.

[1] Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah AH, Appl Environ Microbiol 1993;59:479-484.

348. Magnesium chelatase, subunit ChlI

Magnesium-chelatase is a three-component enzyme that catalyses the insertion of Mg2+ into protoporphyrin IX. This is the first unique step in the synthesis of (bacterio)chlorophyll. Due to this, it is thought that Mg-chelatase has an important role in channeling inter- mediates into the (bacterio)chlorophyll branch in response to conditions suitable for photosynthetic growth. ChlI and BchD have molecular weight between 38-42 kDa.

[1] Walker CJ, Willows RD, Biochem J 1997;327:321-333. [2] Petersen BL, Jensen PE, Gibson LC, Stummann BM, Hunter CN, Henningsen KW, J Bacteriol 1998;180:699-704.

349. Plasmid recombination enzyme (Mob_Pre)

With some plasmids, recombination can occur in a site specific manner that is independent of RecA. In such cases, the recombination event requires another protein called Pre. Pre is a plasmid recombination enzyme. This protein is: also known as Mob (conjugative mobilization).

[1] Priebe SD, Lacks SA, J Bacteriol 1989;171:4778-4784.

350. Monooxygenase

This family includes diverse enzymes that utilise FAD.

[1] Gatti DL, Palfey BA, Lah MS, Entsch B, Massey V, Ballou DP, Ludwig ML, Science 1994;266:110-114.

351. Mov34 family

Members of this family are found in proteasome regulatory subunits, eukaryotic initiation factor 3 (eIF3) subunits and regulators of transcription factors.

[1] Aravind L, Ponting CP, Protein Sci 1998;7:1250-1254. [2] Hershey JW, Asano K, Naranda T, Vornlocher HP, Hanachi P, Merrick WC, Biochimie 1996;78:903-907.

352. Myc amino-terminal region (Myc_N_term)

The myc family belongs to the basic helix-loop-helix leucine zipper class of transcription factors, see HLH. Myc forms a heterodimer with Max, and this complex regulates cell growth through direct activation of genes involved in cell replication [2].

[1] Facchini LM, Penn LZ, FASEB J 1998;12:633-651. [2] Grandori C, Eisenman RN, Trends Biochem Sci 1997;22:177-181.

353. (Metallothio_2) Metallothionein. Members of this family are metallothioneins. These proteins are cysteine rich proteins that bind to heavy metals. Members of this family appear to be closest to Class II metallothioneins, seed metalthio. Number of members: 55

[1] Medline: 98267202. Characterization of gene repertoires at mature stage of citrus fruits through random sequencing and analysis of redundant metallothionein-like genes expressed during fruit development. Moriguchi T, Kita M, Hisada S, Endo-Inagaki T, Omura M; Gene 1998;211:221-227.

354. MAGE family

The MAGE (melanoma antigen-encoding gene) family are expressed in a wide variety of tumors but not in normal cells, with the exception of the male germ cells, placenta, and, possibly,

cells of the developing embryo. The cellular function of this family is unknown.

[1] McCurdy DK, Tai LQ, Nguyen J, Wang Z, Yang HM, Udar N, Naiem F, Concannon P, Gatti RA; Mol Genet Metab 1998;63:3-13.

355. Malic enzymes signature. Malic enzymes, or malate oxidoreductases, catalyze the oxidative decarboxylation of malate into pyruvate important for a wide range of metabolic pathways. There are three related forms of malic enzyme [1,2,3]: - NAD-dependent malic enzyme (EC 1.1.1.38), which uses preferentially NAD and has the ability to decarboxylate oxaloacetate (OAA). It is found in bacteria and insects. - NAD-dependent malic enzyme (EC 1.1.1.39), which uses preferentially NAD and is unable to decarboxylate OAA. It is found in the mitochondrial matrix of plants and is a heterodimer of highly related subunits. - NADP-dependent malic enzyme (EC 1.1.1.40), which has a preference for NADP and has the ability to decarboxylate OAA. This form has been found in fungi, animals and plants. In mammals, there are two isozymes: one, mitochondrial and the other, cytosolic. Plants also have two isozymes: chloroplastic and cytosolic. There are two other proteins which are closely structurally related to malicenzymes: - Escherichia coli protein sfcA, whose function is not yet known but which could be an NAD or NADP-dependent malic enzyme. - Yeast hypothetical protein YKL029c, a probable malic enzyme. There are three well conserved regions in the enzyme sequences. Two of them seem to be involved in binding NAD or NADP. The significance of the third one, located in the central part of the enzymes, is not yet known. This region has been developed as a signature pattern for these enzymes.

Consensus pattern: F-x-[DV]-D-x(2)-G-T-[GSA]-x-[IV]-x-[LIVMA]-[GAST](2)-[LIVMF](2)-

[1] Artus N.N., Edwards G.E. FEBS Lett. 182:225-233(1985).[2] Loeber G., Infante A.A., Maurer-Fogy I., Krystek E., Dworkin M.B. J. Biol. Chem. 266:3016-3021(1991). [3] Long J.J., Wang J.-L., Berry J.O. J. Biol. Chem. 269:2827-2833(1994).

356. (matrixin)

Matrixins cysteine switch (aka peptidase_M10)

Mammalian extracellular matrix metalloproteinases (EC 3.4.24.-), also known as matrixins [1] (see <PDOC00129>), are zinc-dependent enzymes. They are secreted by cells in an inactive form (zymogen) that differs from the mature enzyme by the presence of an N-terminal propeptide. A highly conserved octapeptide is found two residues downstream of the C-terminal end of the propeptide. This region has been shown to be involved in autoinhibition of matrixins [2,3]; a cysteine within the octapeptide chelates the active site zinc ion, thus inhibiting the enzyme. This region has been called the 'cysteine switch' or 'autoinhibitor region'.

A cysteine switch has been found in the following zinc proteases:

- MMP-1 (EC 3.4.24.7) (interstitial collagenase).
- MMP-2 (EC 3.4.24.24) (72 Kd gelatinase).
- MMP-3 (EC 3.4.24.17) (stromelysin-1).
- MMP-7 (EC 3.4.24.23) (matrilysin).
- MMP-8 (EC 3.4.24.34) (neutrophil collagenase).
- MMP-9 (EC 3.4.24.35) (92 Kd gelatinase).
- MMP-10 (EC 3.4.24.22) (stromelysin-2).
- MMP-11 (EC 3.4.24.-) (stromelysin-3).
- MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).
- MMP-13 (EC 3.4.24.-) (collagenase 3).
- MMP-14 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 1).
- MMP-15 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 2).
- MMP-16 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 3).
- Sea urchin hatching enzyme (EC 3.4.24.12) (envelysin) [4].
- Chlamydomonas reinhardtii gamete lytic enzyme (GLE) [5].

Consensus patternP-R-C-[GN]-x-P-[DR]-[LIVSAPKQ] [C chelates the zinc ion] Sequences known to belong to this class detected by the pattern ALL, except for cat MMP-7 and mouse MMP-11.

[1] Woessner J. Jr. FASEB J. 5:2145-2154(1991).

[2] Sanchez-Lopez R., Nicholson R., Gesnel M.C., Matrisian L.M., Breathnach R. J. Biol. Chem. 263:11892-11899(1988).

[3] Park A.J., Matrisian L.M., Kells A.F., Pearson R., Yuan Z., Navre M. J. Biol. Chem. 266:1584-1590(1991).

[4] Lepage T., Gache C. EMBO J. 9:3003-3012(1990).

[5] Kinoshita T., Fukuzawa H., Shimada T., Saito T., Matsuda Y. Proc. Natl. Acad. Sci. U.S.A. 89:4693-4697(1992).

357. Vertebrate metallothioneins signature (metalthio)

Metallothioneins (MT) [1,2,3] are small proteins which bind heavy metals such as zinc, copper, cadmium, nickel, etc., through clusters of thiolate bonds. MT's occur throughout the animal kingdom and are also found in higher plants, fungi and some prokaryotes. On the basis of structural relationships MT's have been subdivided into three classes. Class I includes mammalian MT's as well as MT's from crustacean and molluscs, but with clearly related primary structure. Class II groups together MT's from various species such as sea urchins, fungi, insects and cyanobacteria which display none or only very distant correspondence to class I MT's. Class III MT's are atypical polypeptides containing gamma-glutamylcysteinyl units. Vertebrate class I MT's are proteins of 60 to 68 amino acid residues, 20 of these residues are cysteines that bind to 7 bivalent metal ions. As a signature pattern a region that spans 19 residues and which contains seven of the metal-binding cysteines was chosen, this region is located in the N-terminal section of class-I MT's.

Consensus pattern: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K-

[1] Hamer D.H. Annu. Rev. Biochem. 55:913-951(1986).

[2] Kagi J.H.R., Schaffer A. Biochemistry 27:8509-8515(1988).

[3] Binz P.-A. Thesis, 1996, University of Zurich.

358. Mitochondrial energy transfer proteins signature (mito_ carr)

Different types of substrate carrier proteins involved in energy transfer are found in the inner mitochondrial membrane [1 to 5]. These are: - The ADP,ATP carrier protein (AAC) (ADP/ATP translocase) which exports ATP into the cytosol and imports ADP into the mitochondrial matrix. The sequence of AAC has been obtained from various mammalian, plant and fungal species. - The 2-oxoglutarate/malate carrier protein (OGCP), which exports 2-oxoglutarate into the cytosol and imports malate or other dicarboxylic acids into the mitochondrial matrix. This protein plays an important role in several metabolic processes such as the malate/aspartate and the oxoglutarate/isocitrate shuttles. - The phosphate carrier protein, which transports phosphate groups from the cytosol into the mitochondrial matrix. - The brown fat uncoupling protein (UCP) which dissipates oxidative energy into heat by transporting protons from the cytosol into the mitochondrial matrix. - The tricarboxylate transport protein (or citrate transport protein) which is involved in citrate-H+/malate exchange. It is important for the bioenergetics of hepatic cells as it provides a carbon source for fatty acid and sterol biosyntheses, and NAD for the glycolytic pathway. - The Grave's disease carrier protein (GDC), a protein of unknown function recognized by IgG in patients with active Grave's disease. - Yeast mitochondrial proteins MRS3 and MRS4. The exact function of these proteins is not known. They suppress a mitochondrial splice defect in the first intron of the COB gene and may act as carriers, exerting their suppressor activity by modulating solute concentrations in the mitochondrion. - Yeast mitochondrial FAD carrier protein (gene FLX1). - Yeast protein ACR1 [6], which seems essential for acetyl-CoA synthetase activity. - Yeast protein PET8. - Yeast protein PMT. - Yeast protein RIM2. - Yeast protein YHM1/SHM1. - Yeast protein YMC1. - Yeast protein YMC2. - Yeast hypothetical proteins YBR291c, YEL006w, YER053c, YFR045w, YHR002w, and YIL006w. - Caenorhabditis elegans hypothetical protein K11H3.3.Two other proteins have been found to belong to this family, yet are not localized in the mitochondrial inner membrane: - Maize amyloplast Brittle-1 protein. This protein, found in the endosperm of kernels, could play a role in amyloplast membrane transport. - Candida boidinii peroxisomal membrane protein PMP47 [7]. PMP47 is an integral membrane protein of the peroxisome and it may play a role as a transporter. These proteins all seem to be evolutionary related. Structurally, they consistof three tandem repeats of a domain of approximately one hundred residues. Each of these domains contains two transmembrane regions. As a signature pattern, one of the most conserved regions in the repeated domain was selected, located just after the first transmembrane region.

Consensus pattern: P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]-[QGAIVM]-

[1] Klingenberg M. Trends Biochem. Sci. 15:108-112(1990).

[2] Walker J.E. Curr. Opin. Struct. Biol. 2:519-526(1992).

[3] Kuan J., Saier M.H. Jr. CRC Crit. Rev. Biochem. 28:209-233(1993).

[4] Kuan J., Saier M.H. Jr. Res. Microbiol. 144:671-672(1993).

[5] Nelson D.R., Lawson J.E., Klingenberg M., Douglas M.G. J. Mol. Biol. 230:1159-1170(1993).

[6] Palmieri F. FEBS Lett. 346:48-54(1994).

[7] Jank B., Habermann B., Schweyen R.J., Link T.A. Trends Biochem. Sci. 18:427-428(1993).

359. Prokaryotic molybdopterin oxidoreductases signatures (molybdopterin)

A number of different prokaryotic oxidoreductases that require and bind amolybdopterin cofactor have been shown [1,2,3] to share a number of regions of sequence similarity. These enzymes are: - Escherichia coli respiratory nitrate reductase (EC 1.7.99.4). This enzyme complex allows the bacteria to use nitrate as an electron acceptor during anaerobic growth. The enzyme is composed of three different chains: alpha, beta and gamma. The alpha chain (gene narG) is the molybdopterin-binding subunit. Escherichia coli encodes for a second, closely related, nitrate reductase complex which also contains a molybdopterin-binding alpha chain (gene narZ). - Escherichia coli anaerobic dimethyl sulfoxide reductase (DMSO reductase). DMSO reductase is the terminal reductase during anaerobic growth on various sulfoxide and N-oxide compounds. DMSO reductase is composed of three chains: A, B and C. The A chain (gene dmsA) binds molybdopterin. - Escherichia coli biotin sulfoxide reductases (genes bisC and bisZ). This enzyme reduces a spontaneous oxidation product of biotin, BDS, back to biotin. It may serve as a scavenger, allowing the cell to use biotin sulfoxide as a biotin source. - Methanobacterium formicicum formate dehydrogenase (EC 1.2.1.2). The alpha chain (gene fdhA) of this dimeric enzyme binds a molybdopterin cofactor. - Escherichia coli formate dehydrogenases -H (gene fdhF), -N (gene fdnG) and -O (gene fdoG). These enzymes are responsible for the oxidation of formate to carbon dioxide. In addition to molybdopterin, the alpha (catalytic) subunit also contains an active site, selenocysteine. - Wolinella succinogenes polysulfide reductase chain. This enzyme is a component of the phosphorylative electron transport system with polysulfide as the terminal

acceptor. It is composed of three chains: A, B and C. The A chain (gene psrA) binds molybdopterin. - Salmonella typhimurium thiosulfate reductase (gene phsA). - Escherichia coli trimethylamine-N-oxide reductase (EC 1.6.6.9) (gene torA) [4]. - Nitrate reductase (EC 1.7.99.4) from Klebsiella pneumoniae (gene nasA), Alcaligenes eutrophus, Escherichia coli, Rhodobacter sphaeroides, Thiosphaera pantotropha (gene napA), and Synechococcus PCC 7942 (gene narB).These proteins range from 715 amino acids (fdhF) to 1246 amino acids (narZ) insize. Three signature patterns for these enzymes were derived. The first is based on a conserved region in the N-terminal section and contains two cysteine residues perhaps involved in binding the molybdopterin cofactor. It should be noted that this region is not present in bisC. The second pattern is derived from a conserved region located in the central part of these enzymes.

Consensus pattern: [STAN]-x-[CH]-x(2,3)-C-[STAG]-[GSTVMF]-x-C-x-[LIVMFYW]-x-[LIVMA]-x(3,4)-[DENQKHT]-

Consensus pattern: [STA]-x-[STAC](2)-x(2)-[STA]-D-[LIVMY](2)-L-P-x-[STAC](2)- x(2)-E-

Consensus pattern: A-x(3)-[GDT]-I-x-[DNQTK]-x-[DEA]-x-[LIVM]-x-[LIVMC]-x- [NS]-x(2)-[GS]-x(5)-A-x-[LIVM]-[ST]-

[1] Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C. Biochim. Biophys. Acta 1057:157-185(1991).

[2] Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H. Mol. Microbiol. 2:785-795(1988).

[3] Trieber C.A., Rothery R.A., Weiner J.H. J. Biol. Chem. 269:7103-7109(1994).

[4] Mejean V., Lobbi-Nivol C., Lepelletier M., Giordano G., Chippaux M., Pascal M.-C. Mol. Microbiol. 11:1169-1179(1994).

360. Bacterial mutT domain signature

The bacterial mutT protein is involved in the GO system [1] responsible for removing an oxidatively damaged form of guanine (8-hydroxyguanine or7,8-dihydro-8-oxoguanine) from DNA and the nucleotide pool. 8-oxo-dGTP is inserted opposite to dA and dC residues of template DNA with almost equal efficiency thus leading to A.T to G.C transversions. MutT specifically degrades 8-oxo-dGTP to the monophosphate with the concomitant release of pyrophosphate. MutT is a small protein of about 12 to 15 Kd. It has been shown [2,3] that a

region of about 40 amino acid residues, which is found in the N-terminal part of mutT, can also be found in a variety of other prokaryotic, viral, and eukaryotic proteins. These proteins are:

- Streptomyces pneumoniae mutX.
- A mutT homolog from plasmid pSAM2 of Streptomyces ambofaciens.
- Bartonella bacilliformis invasion protein A (gene invA).
- Escherichia coli dATP pyrophosphohydrolase.
- Protein D250 from African swine fever viruses.
- Proteins D9 and D10 from a variety of poxviruses.
- Mammalian 7,8-dihydro-8-oxoguanine triphosphatase (EC 3.1.6.-) [4].
- Mammalian diadenosine 5',5'''-P1,P4-tetraphosphate asymmetrical hydrolase (Ap4Aase) (EC 3.6.1.17) [5], which cleaves A-5'-PPPP-5'A to yield AMP and ATP.
- A protein encoded on the antisense RNA of the basic fibroblast growth factor gene in higher vertebrates.
- Yeast protein YSA1.
- Escherichia coli hypothetical protein yfaO.
- Escherichia coli hypothetical protein ygdU and HI0901, the corresponding Haemophilus influenzae protein.
- Escherichia coli hypothetical protein yjaD and HI0432, the corresponding Haemophilus influenzae protein.
- Escherichia coli hypothetical protein yrfE.
- Bacillus subtilis hypothetical protein yqkG.
- Bacillus subtilis hypothetical protein yzgD.
- Yeast hypothetical protein YGL067w.

It is proposed [2] that the conserved domain could be involved in the active center of a family of pyrophosphate-releasing NTPases. As a signature pattern the core region of the domain was selected; it contains four conserved glutamate residues.

Consensus pattern: G-x(5)-E-x(4)-[STAGC]-[LIVMAC]-x-R-E-[LIVMFT]-x-E-E-

[1] Michaels M.L., Miller J.H. J. Bacteriol. 174:6321-6325(1992).
[2] Koonin E.V. Nucleic Acids Res. 21:4847-4847(1993).

[3] Mejean V., Salles C., Bullions M.J., Bessman M.J., Claverys J.-P. Mol. Microbiol. 11:323-330(1994).

[4] Sakumi K., Furuichi M., Tsuzuki T., Kakuma T., Kawabata S., Maki H., Sekiguchi M. J. Biol. Chem. 268:23524-23530(1993).

[5] Thorne N.M.H., Hankin S., Wilkinson M.C., Nunez C., Barraclough R., McLennan A.G. Biochem. J. 311:717-721(1995).

361. Myb DNA-binding domain repeat signatures

The retroviral oncogene v-myb , and its cellular counterpart c-myb, encodenuclear DNA-binding proteins that specifically recognize the sequence YAAC(G/T)G [1]. The myb family also includes the following proteins: - Drosophila D-myb [2]. - Vertebrate myb-like proteins A-myb and B-myb [3]. - Maize C1 protein, a trans-acting factor which controls the expression of genes involved in anthocyanin biosynthesis. - Maize P protein [4], a trans-acting factor which regulates the biosynthetic pathway of a flavonoid-derived pigment in certain floral tissues. - Arabidopsis thaliana protein GL1 [5], required for the initiation of differentiation of leaf hair cells (trichomes). - A number of myb/c1-related proteins in maize and barley, whose roles are not yet known [4]. - Yeast BAS1 [7], a transcriptional activator for the HIS4 gene. - Yeast REB1 [8], which recognizes sites within both the enhancer and the promoter of rRNA transcription, as well as upstream of many genes transcribed by RNA polymerase II. - Fission yeast cdc5, a possible transcription factor whose activity is required for cell cycle progression and growth during G2. - Fission yeast myb1, which regulates telomere length and function. - Yeast hypothetical protein YMR213w.One of the most conserved regions in all of these proteins is a domain of 160amino acids. It consists of three tandem repeats of 51 to 53 amino acids. In myb, this repeat region has been shown [9] to be involved in DNA-binding. The major part of the first repeat is missing in retroviral v-myb sequences and in plant myb-related proteins. Yeast REB1 differs from the other proteins in this family in having a single myb-like domain. As shown in the following schematic representation, two signature patterns for myb-like domains were developed; the first is located in the N-terminal section, the second spans the C-terminal extremity of the domain.

xxxxxxxxxWxxxEDxxxxxxxxxxxxxxWxxIxxxxxxRxxxxxxxxWxxxxx *********
*************************'*' : Position of the patterns.

Consensus pattern: W-[ST]-x(2)-E-[DE]-x(2)-[LIV]-

Consensus pattern: W-x(2)-[LI]-[SAG]-x(4,5)-R-x(8)-[YW]-x(3)-[LIVM]-

Note: this pattern detects the three copies of the domain in myb, d-myb, A-myb and B-myb; the second of the two complete copies of plant myb-related proteins, and the last two copies of yeast BAS1

[1] Biednkapp H., Borgmeyer U., Sippel A.E., Klempnauer K.-H. Nature 335:835-837(1988).

[2] Peters C.W.B., Sippel A.E., Vingron M., Klempnauer K.-H. EMBO J. 6:3085-3090(1987).

[3] Nomura N., Takahashi M., Matsui M., Ishii S., Date T., Sasamoto S., Ishizaki R. Nucleic Acids Res. 16:11075-11090(1988).

[4] Grotewold E., Athma P., Peterson T. Proc. Natl. Acad. Sci. U.S.A. 88:4587-4591(1991).

[5] Oppenheimer D.G., Herman P.L., Sivakumaran S., Esch J., Marks M.D. Cell 67:483-493(1991).

[6] Marocco A., Wissenbach M., Becker D., Paz-Ares J., Saedler H., Salamini F., Rohde W. Mol. Gen. Genet. 216:183-187(1989).

[7] Tice-Baldwin K., Fink G.R., Arndt K.T. Science 246:931-935(1989).

[8] Ju Q., Morrow B.E., Warner J.R. Mol. Cell. Biol. 10:5226-5234(1990).

[9] Klempnauer K.-H., Sippel A.E. EMBO J. 6:2719-2725(1987).

362. NAD-dependent glycerol-3-phosphate dehydrogenase signature

NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3- phosphate. It is a eukaryotic cytosolic homodimeric protein of about 40 Kd. As a signature pattern a glycine-rich region that is probably [1] involved in NAD-binding was selected.

Consensus pattern: G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x- [DE]-G-[LIVM]-x-[LIVMFYW]-G-x-N-

[1] Otto J., Argos P., Rossmann M.G. Eur. J. Biochem. 109:325-330(1980).

363. Nucleosome assembly protein (NAP)

It is thought that NAPs may be involved in regulating gene expression as a result of histone accessibility [1].

[1] Rodriguez P, Munroe D, Prawitt D, Chu LL, Bric E, Kim J, Reid LH, Davies C, Nakagama H, Loebbert R, Winterpacht A, Petruzzi MJ, Higgins MJ, Nowak N, Evans G, Shows T, Weissman BE, Zabel B, Housman DE, Pelletier J, Genomics 1997;44:253-265. [2] Schnieders F, Dork T, Arnemann J, Vogel T, Werner M, Schmidtke J; Hum Mol Genet 1996;5:1801-1807.

364. NB-ARC domain

van der Biezen EA, Jones JD, Curr Biol 1998;8:226-227.

365. Nucleoside diphosphate kinases active site

Nucleoside diphosphate kinases (EC <u>2.7.4.6</u>) (NDK) [1] are enzymes required for the synthesis of nucleoside triphosphates (NTP) other than ATP. They provide NTPs for nucleic acid synthesis, CTP for lipid synthesis, UTP for polysaccharide synthesis and GTP for protein elongation, signal transduction and microtubule polymerization. In eukaryotes, there seems to be a small family of NDK isozymes each of which acts in a different subcellular compartment and/or has a distinct biological function. Eukaryotic NDK isozymes are hexamers of two highly related chains (Aand B) [2]. By random association (A6, A5B...AB5, B6), these two kinds of chain form isoenzymes differing in their isoelectric point. NDK are proteins of 17 Kd that act via a ping-pong mechanism in which a histidine residue is phosphorylated, by transfer of the terminal phosphate group from ATP. In the presence of magnesium, the phosphoenzyme can transfer its phosphate group to any NDP, to produce an NTP.NDK isozymes have been sequenced from prokaryotic and eukaryotic sources. It has also been shown [3] that the Drosophila awd (abnormal wing discs) protein, is a microtubule-associated NDK. Mammalian NDK is also known as metastasis inhibition factor nm23.The sequence of NDK has been highly conserved through evolution. There is a single histidine residue conserved in all known NDK isozymes, which is involved in the catalytic mechanism [2]. Our signature pattern contains this residue.

Consensus pattern: N-x(2)-H-[GA]-S-D-[SA]-[LIVMPKNE] [H is the putative active site residue]-

[1] Parks R., Agarwal R. (In) The Enzymes (3rd edition) 8:307-334(1973).
[2] Gilles A.-M., Presecan E., Vonica A., Lascu I. J. Biol. Chem. 266:8784-8789(1991).
[3] Biggs J., Hersperger E., Steeg P.S., Liotta L.A., Shearn A. Cell 63:933-940(1990).

366. Nitrite and sulfite reductases iron-sulfur/siroheme-binding site (NIR_SIR)
Nitrite reductases (NiR) [1] catalyze the reduction of nitrite into ammonium, the second step in the assimilation of nitrate. There are two types of NiR: the higher plant chloroplastic form of NiR (EC 1.7.7.1) is a monomeric protein that uses reduced ferredoxin as the electron donor; while fungal and bacterial NiR (EC 1.6.6.4) are homodimeric proteins that uses NAD(P)H as the electron donor. Both forms of NiR contain a siroheme-Fe and iron-sulfur centers. Sulfite reductase (NADPH) (EC 1.8.1.2) (SIR) [2] is the bacterial enzyme that catalyzes the reduction of sulfite to sulfide. SIR is an oligomeric enzyme with a subunit composition of alpha(8)-beta(4), the alpha component is a flavoprotein (SIR-FP), while the beta component is a siroheme, iron-sulfurprotein (SIR-HP).Sulfite reductase (ferredoxin) (EC 1.8.7.1) [3] is a cyanobacterial and plant monomeric enzyme that also catalyzes the reduction of sulfite to sulfide. Anaerobic sulfite reductase (EC 1.8.1.-) (ASR) [4], a bacterial enzyme that catalyzes the NADH-dependent reduction of sulfite to sulfide. ASR is an oligomeric enzyme composed of three different subunits. The C component (geneasrC) seems to be a siroheme, iron-sulfur protein. These enzymes share a region of sequence similarity in their C-terminal half; this region which spans about 80 amino acids includes four conserved cysteine residues. Two of the Cys are grouped together at the beginning of the domain, and the two others are grouped in the middle of the domain. The cysteines are involved in the binding of the iron-sulfur center; the last one also binds the siroheme group [2]. A signature pattern from the region around the second cluster of cysteines was derived.

Consensus pattern: [STV]-G-C-x(3)-C-x(6)-[DE]-[LIVMF]-[GAT]-[LIVMF] [The two C's are ison-sulfur ligands]-

[1] Campbell W.H., Kinghorn J.R. Trends Biochem. Sci. 15:315-319(1990).
[2] Crane B.R., Siegel L.M., Getzoff E.D. Science 270:59-67(1995).
[3] Gisselmann G., Klausmeier P., Schwenn J.D. Biochim. Biophys. Acta 1144:102-106(1993).

[4] Huang C.J., Barrett E.L. J. Bacteriol. 173:1544-1553(1991).

367. (NMT) Myristoyl-CoA:protein N-myristoyltransferase signatures. Myristoyl-CoA: protein N-myristoyltransferase (EC 2.3.1.97) (Nmt) [1] is the enzyme responsible for transferring a myristate group on the N-terminal glycine of a number of cellular eukaryotic and viral proteins. Nmt is a monomeric protein of about 50 to 60 Kd whose sequence appears to be well conserved. Two highly conserved regions have been developed as signature patterns. The first one is located in the central section, the second in the C-terminal part.

Consensus pattern: E-I-N-F-L-C-x-H-K-
Consensus pattern: K-F-G-x-G-D-G-

[1] Rudnick D.A., McWherter C.A., Gokel G.W., Gordon J.I. Adv. Enzymol. 67:375-430(1993).

368. ADP-glucose pyrophosphorylase signatures (NTP_transferase)

ADP-glucose pyrophosphorylase (glucose-1-phosphate adenylyltransferase) [1,2](EC 2.7.7.27) catalyzes a very important step in the biosynthesis of alpha 1,4-glucans (glycogen or starch) in bacteria and plants: synthesis of the activated glucosyl donor, ADP-glucose, from glucose-1-phosphate and ATP.ADP-glucose pyrophosphorylase is a tetrameric allosterically regulated enzyme. It is a homotetramer in bacteria while in plant chloroplasts and amyloplasts, it is a heterotetramer of two different, yet evolutionary related, subunits. There are a number of conserved regions in the sequence of bacterial and plant ADP-glucose pyrophosphorylase subunits. Three of these regions were selected as signature patterns. The first two are N-terminal and have been proposed to be part of the allosteric and/or substrate-binding sites in the Escherichia coli enzyme (gene glgC). The third pattern corresponds to a conserved region in the central part of the enzymes.

Consensus pattern: [AG]-G-G-x-G-[STK]-x-L-x(2)-L-[TA]-x(3)-A-x-P-A-[LV] -
Consensus pattern: W-[FY]-x-G-[ST]-A-[DNSH]-[AS]-[LIVMFYW]-
Consensus pattern: [APV]-[GS]-M-G-[LIVMN]-Y-[IVC]-[LIVMFY]-x(2)-[DENPHK] -

[1] Nakata P.A., Greene T.W., Anderson J.M., Smith-White B.J., Okita T.W., Preiss J. Plant Mol. Biol. 17:1089-1093(1991).

[2] Preiss J., Ball K., Hutney J., Smith-White B.J., Li. L., Okitsa T.W. Pure Appl. Chem. 63:535-544(1991).

369. Sodium/hydrogen exchanger family

Na/H antiporters are key transporters in maintaining the pH of actively metabolizing cells. The molecular mechanisms of antiport are unclear.

These antiporters contain 10-12 transmembrane regions (M) at the amino-terminus and a large cytoplasmic region at the carboxyl terminus. The transmembrane regions M3-M12 share identity with other members of the family. The M6 and M7 regions are highly conserved. Thus, this is thought to be the region that is involved in the transport of sodium and hydrogen ions. The cytoplasmic region has little similarity throughout the family.

[1] Dibrov P, Fliegel L; FEBS Lett 1998;424:1-5. [2] Orlowski J, Grinstein S; J Biol Chem 1997;272:22373-22376.[3] Numata M, Petrecca K, Lake N, Orlowski J; J Biol Chem 1998;273:6951-6959.

370. Sodium:sulfate symporter family signature (Na_sulph_symp)

Integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) canbe grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families currently consists of the following proteins: - Mammalian sodium/sulfate cotransporter [1]. - Mammalian renal sodium/dicarboxylate cotransporter [2], which transports succinate and citrate. - Mammalian intestinal sodium/dicarboxylate cotransporter. - Chlamydomonas reinhardtii putative sulfur deprivation response regulator SAC1 [3]. - Caenorhabditis elegans hypothetical proteins B0285.6, F31F6.6, K08E5.2 and R107.1. - Escherichia coli hypothetical protein yfbS. - Haemophilus influenzae hypothetical protein HI0608. - Synechocystis strain PCC 6803 hypothetical protein sll0640. - Methanococcus jannaschii hypothetical protein MJ0672.These transporters are proteins of from 430 to 620 amino acids which are highly

hydrophobic and which probably contain about 12 transmembrane regions. As a signature pattern, a conserved region was selected which is located in or near the penultimate transmembrane region.

Consensus pattern: [STACP]-S-x(2)-F-x(2)-P-[LIVM]-[GSA]-x(3)-N-x-[LIVM]-V-

[1] Markovich D., Forgo J., Stange G., Biber J., Murer H. Proc. Natl. Acad. Sci. U.S.A. 90:8073-8077(1993).

[2] Pajor A.M. Am. J. Physiol. 270:642-648(1996).

[3] Davies J.P., Yildiz F.H., Grossman A. EMBO J. 15:2150-2159(1996).

371. NifU-like domain

This is an alignment of the carboxy-terminal domain. This is the only common region between the NifU protein from nitrogen-fixing bacteria and rhodobacterial species. The biochemical function of NifU is unknown [1].

Ouzounis C, Bork P, Sander C, Trends Biochem Sci 1994;19:199-200.

372. Nitrilases / cyanide hydratase signatures

Nitrilases (EC 3.5.5.1) are enzymes that convert nitriles into their corresponding acids and ammonia. They are widespread in microbes as well as in plants where they convert indole-3-acetonitrile to the hormone indole-3-acetic acid. A conserved cysteine has been shown [1,2] to be essential for enzyme activity; it seems to be involved in a nucleophilic attack on the nitrile carbon atom. Cyanide hydratase (EC 4.2.1.66) converts HCN to formamide. In phytopathogenic fungi, it is used to avoid the toxic effect of cyanide released by wounded plants [3]. The sequence of cyanide hydrolase is evolutionary related to that of nitrilases. Yeast hypothetical proteins YIL164c and YIL165c also belong to this family. As signature patterns for these enzymes, two conserved regions were selected. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central section.

Consensus pattern: G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P-

Consensus pattern: G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR] [C is the active site residue]-

[1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993).
[2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yamada H. J. Biol. Chem. 267:20746-20751(1992).
[3] Wang P., Vanetten H.D. Biochem. Biophys. Res. Commun. 187:1048-1054(1992).

373. NusB family

The NusB protein is involved in the regulation of rRNA biosynthesis by transcriptional antitermination.

Huenges M, Rolz C, Gschwind R, Peteranderl R, Berglechner F, Richter G, Bacher A, Kessler H,Gemmecker G, EMBO J 1998;17:4092-4100.

374. (Neur Chan) Neurotransmitter-gated ion-channels signature

Neurotransmitter-gated ion-channels [1,2,3,4] provide the molecular basis for rapid signal transmission at chemical synapses. They are post-synapticoligomeric transmembrane complexes that transiently form a ionic channel upon the binding of a specific neurotransmitter. Presently, the sequence of subunits from five types of neurotransmitter-gated receptors are known: - The nicotinic acetylcholine receptor (AchR), an excitatory cation channel. In the motor endplates of vertebrates, it is composed of four different subunits (alpha, beta, gamma and delta or epsilon) with a molar stoichiometry of 2:1:1:1. In neurones, the AchR receptor is composed of two different types of subunits: alpha and non-alpha (also called beta). Nicotinic AchRs are also found in invertebrates. - The glycine receptor, an inhibitory chloride ion channel. The glycine receptor is a pentamer composed of two different subunits (alpha and beta). - The gamma-aminobutyric-acid (GABA) receptor, which is also an inhibitory chloride ion channel. The quaternary structure of the GABA receptor is complex; at least four classes of subunits are known to exist (alpha, beta, gamma, and delta) and there are many variants in each class (for example: six variants of the alpha class have already been sequenced). - The serotonin 5HT3 receptor. Serotonin is a biogenic hormone that functions as a neurotransmitter, a hormone and a mitogen. There are seven major groups

of serotonin receptors; six of these groups (5HT1, 5HT2, and 5HT4 to 5HT7) transduce extracellular signal by activating G proteins, while 5HT3 is a ligand-gated cation-specific ion channel which, when activated causes fast, depolarizing responses in neurons. - The glutamate receptor, an excitatory cation channel. Glutamate is the main excitatory neurotransmitter in the brain. At least three different types of glutamate receptors have been described and are named according to their selective agonists (kainate, N-methyl-D-aspartate (NMDA) and quisqualate).All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related. They are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions which form the ionic channel, followed by an intracellular region of variable length. A fourth hydrophobic region is found at the C-terminal of the sequence. The sequence of subunits from the AchR, GABA, 5HT3, and Gly receptors are clearly evolutionary related and share many regions of sequence similarities. These sequence similarities are either absent or very weak in the Glu receptors. In the N-terminal extracellular domain of AchR/GABA/5HT3/Gly receptors, there are two conserved cysteine residues, which, in AchR, have been shown to form a disulfide bond essential to the tertiary structure of the receptor. A number of amino acids between the two disulfide-bonded cysteines are also conserved. Therefore this region was used as a signature pattern for this subclass of proteins.

Consensus pattern: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C [The two C's are linked by a disulfide bond]-

[1] Stroud R.M., McCarthy M.P., Shuster M. Biochemistry 29:11009-11023(1990).
[2] Betz H. Neuron 5:383-392(1990).
[3] Dingledine R., Myers S.J., Nicholas R.A. FASEB J. 4:2632-2645(1990).
[4] Barnard E.A. Trends Biochem. Sci. 17:368-374(1992).

375. Orotidine 5'-phosphate decarboxylase active site

Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMPdecase) [1,2] catalyzes the last step in the de novo biosynthesis of pyrimidines, the decarboxylation of OMP into UMP. In higher eukaryotes OMPdecase is part, with orotatephosphoribosyltransferase, of a bifunctional enzyme, while the prokaryotic and fungal OMPdecases are monofunctional protein. Some parts of the sequence of OMPdecase are well conserved across species. The best conserved

region is located in the N-terminal half of OMPdecases and is centered around a lysine residue which is essential for the catalytic function of the enzyme. This region has been developed as a signature pattern.

Consensus pattern: [LIVMFTA]-[LIVMF]-x-D-x-K-x(2)-D-I-[GP]-x-T-[LIVMTA] [K is the active site residue]-

[1] Jacquet M., Guilbaud R., Garreau H. Mol. Gen. Genet. 211:441-445(1988).
[2] Kimsey H.H., Kaiser D. J. Biol. Chem. 267:819-824(1992).

376. ATP synthase delta (OSCP) subunit signature

ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1).

One of the subunits of the ATPase complex, known as subunit delta in bacteria and chloroplasts or the Oligomycin Sensitivity Conferral Protein (OSCP) in mitochondria, seems to be part of the stalk that links CF(0) to CF(1). It either transmits conformational changes from CF(0) into CF(1) or is involved in proton conduction [3].

The different delta/OSCP subunits are proteins of approximately 200 amino-acid residues - once the transit peptide has been removed in the chloroplast and mitochondrial forms - which show only moderate sequence homology.

The signature pattern used to detect ATPase delta/OSCP subunits is based on a conserved region in the C-terminal section of these proteins.

Consensus pattern: [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x-[GSA]-G-[LIVMFYGA]-x-[LIVM]-[KRHENQ]-x-[GSEN]

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).
[2] Senior A.E. Physiol. Rev. 68:177-231(1988).
[3] Engelbrecht S., Junge W. Biochim. Biophys. Acta 1015:379-390(1990).

377. Aspartate and ornithine carbamoyltransferases signature

Aspartate carbamoyltransferase (EC 2.1.3.2) (ATCase) catalyzes the conversion of aspartate and carbamoyl phosphate to carbamoylaspartate, the second step in the de novo biosynthesis of pyrimidine nucleotides [1]. In prokaryotes ATCase consists of two subunits: a catalytic chain (gene pyrB) and a regulatory chain (gene pyrI), while in eukaryotes it is a domain in a multi-functional enzyme (called URA2 in yeast, rudimentary in Drosophila, and CAD in mammals [2]) that also catalyzes other steps of the biosynthesis of pyrimidines.

Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase) catalyzes the conversion of ornithine and carbamoyl phosphate to citrulline. In mammals this enzyme participates in the urea cycle [3] and is located in the mitochondrial matrix. In prokaryotes and eukaryotic microorganisms it is involved in the biosynthesis of arginine. In some bacterial species it is also involved in the degradation of arginine [4] (the arginine deaminase pathway).

It has been shown [5] that these two enzymes are evolutionary related. The predicted secondary structure of both enzymes are similar and there are some regions of sequence similarities. One of these regions includes three residues which have been shown, by crystallographic studies [6], to be implicated in binding the phosphoryl group of carbamoyl phosphate. This region was selected as a signature for these enzymes.

Consensus pattern: F-x-[EK]-x-S-[GT]-R-T[S, R, and the 2nd T bind carbamoyl phosphate]

-Note: the residue in position 3 of the pattern allows to distinguish between an ATCase (Glu) and an OTCase (Lys).

[1] Lerner C.G., Switzer R.L. J. Biol. Chem. 261:11156-11165(1986).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[3] Takiguchi M., Matsubasa T., Amaya Y., Mori M. BioEssays 10:163-166(1989).

[4] Baur H., Stalon V., Falmagne P., Luethi E., Haas D. Eur. J. Biochem. 166:111-117(1987).

[5] Houghton J.E., Bencini D.A., O'Donovan G.A., Wild J.R. Proc. Natl. Acad. Sci. U.S.A. 81:4864-4868(1981).

[6] Ke H.-M., Honzatko R.B., Lipscomb W.N. Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).

378. Oleosins signature

Oleosins [1] are the proteinaceous components of plants' lipid storage bodies called oil bodies. Oil bodies are small droplets (0.2 to 1.5 mu-m in diameter) containing mostly triacylglycerol that are surrounded by a phospholipid/ oleosin annulus. Oleosins may have a structural role in stabilizing the lipid body during dessication of the seed, by preventing coalescence of the oil. They may also provide recognition signals for specific lipase anchorage in lipolysis during seedling growth. Oleosins are found in the monolayer lipid/ water interface of oil bodies and probably interact with both the lipid and phospholipid moieties.

Oleosins are proteins of 16 Kd to 24 Kd and are composed of three domains: an N-terminal hydrophilic region of variable length (from 30 to 60 residues); a central hydrophobic domain of about 70 residues and a C-terminal amphipathic region of variable length (from 60 to 100 residues). The central hydrophobic domain is proposed to be made up of beta-strand structure and to interact with the lipids [2]. It is the only domain whose sequence is conserved and therefore a section from that domain was selected as a signature pattern.

Consensus pattern: [AG]-[ST]-x(2)-[AG]-x(2)-[LIVM]-[SAD]-T-P-[LIVMF](4)-F-S-P-[LIVM](3)-P-A

[1] Murphy D.J., Keen J.N., O'Sullivan J.N., Au D.M.Y., Edwards E.-W., Jackson P.J., Cummins I., Gibbons T., Shaw C.H., Ryan A.J. Biochim. Biophys. Acta 1088:86-94(1991).

[2] Tzen J.T.C., Lie G.C., Huang A.H.C. J. Biol. Chem. 267:15626-15634(1992).

379. (Orbi VP5) Orbivirus outer capsid protein VP5

This paper shows the location of the different capsid proteins and their relation to each other.

[1] Schoehn G, Moss SR, Nuttall PA, Hewat EA; Virology 1997;235:191-200.

380. Orn/DAP/Arg decarboxylases family 2 signatures

Pyridoxal-dependent decarboxylases acting on ornithine, lysine, arginine and related substrates can be classified into two different families on the basis of sequence similarities [1,2,3]. The second family consists of:

- Eukaryotic ornithine decarboxylase (EC 4.1.1.17) (ODC). ODC catalyzes the transformation of ornithine into putrescine.
- Prokaryotic diaminopimelic acid decarboxylase (EC 4.1.1.20) (DAPDC). DAPDC catalyzes the conversion of diaminopimelic acid into lysine; the last step in the biosynthesis of lysine.
- Pseudomonas syringae pv. tabaci protein tabA. tabA is probably involved in the biosynthesis of tabtoxin and is highly similar to DAPDC.
- Bacterial and plant biosynthetic arginine decarboxylase (EC 4.1.1.19) (ADC). ADC catalyzes the transformation of arginine into agmatine, the first step in the biosynthesis of putrescine from arginine.

The above proteins, while most probably evolutionary related, do not share extensive regions of sequence similarities. Two of the conserved regions were selected as signature patterns. The first pattern contains a conserved lysine residue which is known, in mouse ODC [4], to be the site of attachment of the pyridoxal-phosphate group. The second pattern contains a stretch of three consecutive glycine residues and has been proposed to be part of a substrate-binding region [5].

These enzymes are collectively known as group IV decarboxylases [3].

Consensus pattern: [FY]-[PA]-x-K-[SACV]-[NHCLFW]-x(4)-[LIVMF]-[LIVMTA]-x(2)-[LIVMA]-x(3)-[GTE] [K is the pyridoxal-P attachment site]

Consensus pattern: [GS]-x(2,6)-[LIVMSCP]-x(2)-[LIVMF]-[DNS]-[LIVMCA]-G-G-G-[LIVMFY]-[GSTPCEQ]

[1] Bairoch A. Unpublished observations (1993).
[2] Martin C., Cami B., Yeh P., Stragier P., Parsot C., Patte J.-C. Mol. Biol. Evol. 5:549-559(1988).
[3] Sandmeier E., Hale T.I., Christen P. Eur. J. Biochem. 221:997-1002(1994).
[4] Poulin R., Lu L., Ackermann B., Bey P., Pegg A.E. J. Biol. Chem. 267:150-158(1992).
[5] Moore R.C., Boyle S.M. J. Bacteriol. 172:4631-4640(1990).

381. Osteopontin signature

Osteopontin is an acidic phosphorylated glycoprotein of about 40 Kd which is abundant in the mineral matrix of bones and which binds tightly to hydroxyapatite [1,2,3]. It is suggested that osteopontin might function as a cell attachment factor and could play a key role in the adhesion of osteoclasts to the mineral matrix of bone.

Osteopontin-K is a kidney protein which is highly similar to osteopontin and probably also involved in cell-adhesion.

As a signature pattern a highly conserved region located at the N-terminal extremity of the mature protein was selected.

Consensus pattern: [KQ]-x-[TA]-x(2)-[GA]-S-S-E-E-K

[1] Butler W.T. Connect. Tissue Res. 23:123-36(1989).
[2] Gorski J.P. Calcif. Tissue Int. 50:391-396(1992).
[3] Denhardt D.T., Guo X. FASEB J. 7:1475-1482(1993).

382. Oxysterol-binding protein family signature

A number of eukaryotic proteins that seem to be involved with sterol synthesis and/or its regulation have been found [1] to be evolutionary related:

- Mammalian oxysterol-binding protein (OSBP). A protein of about 800 amino-acid residues that binds a variety of oxysterols: oxygenated derivatives of cholesterol. OSBP seems to play a complex role in the regulation of sterol metabolism.
- Yeast proteins HES1 and KES1; highly related proteins of 434 residues that

seem to play a role in ergosterol synthesis.

- Yeast OSH1, a protein of 859 residues that also plays a role in ergosterol synthesis. - Yeast hypothetical protein YHR001w (437 residues).
- Yeast hypothetical protein YHR073w (996 residues).
- Yeast hypothetical protein YKR003w (448 residues).

All these proteins contain a moderately conserved domain of about 250 residues located in the C-terminal half of OBSP, OSH1 and YHR073w and in the central section of the other proteins. As a signature pattern, the best conserved part was selected of this domain, a region that contains a conserved pentapeptide.

Consensus pattern: E-[KQ]-x-S-H-[HR]-P-P-x-[STACF]-A

[1] Jiang B., Brown J.L., Sheraton J., Fortin N., Bussey H. Yeast 10:341-353(1994).

383. FMN oxidoreductase

384. Oxidoreductase FAD/NAD-binding domain

Number of members: 250

[1]
Medline: 92084635
The sequence of squash NADH:nitrate reductase and its relationship to the sequences of other flavoprotein oxidoreductases. A family of flavoprotein pyridine nucleotide cytochrome reductases.
Hyde GE, Crawford NM, Campbell W;
J Biol Chem 1991;266:23542-23547.
[2]Medline: 95111952
Crystal structure of the FAD-containing fragment of corn nitrate reductase at 2.5 A resolution: relationship to other flavoprotein reductases.
Lu G, Campbell WH, Schneider G, Lindqvist Y;

Structure 1994;2:809-821.

385. (oxidored molyb) Eukaryotic molybdopterin oxidoreductases signature

A number of different eukaryotic oxidoreductases that require and bind a molybdopterin cofactor have been shown [1] to share a few regions of sequence similarity. These enzymes are:

- Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of xanthine to uric acid with the concomitant reduction of NAD. Structurally, this enzyme of about 1300 amino acids consists of at least three distinct domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain (see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mo-pterin domain.
- Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its sequence and domain structure.
- Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate to nitrite. Structurally, this enzyme of about 900 amino acids consists of an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome reductase domain.
- Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to sulfate. Structurally, this enzyme of about 460 amino acids consists of an N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.

There are a few conserved regions in the sequence of the molybdopterin-binding domain of these enzymes. The pattern used to detect these proteins is based on one of them. It contains a cysteine residue which could be involved in binding the molybdopterin cofactor.

Consensus pattern: [GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-x(2)-[DE]

[1] Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C. Biochim. Biophys. Acta 1057:157-185(1991).

386. (Oxidored q1) NADH-Ubiquinone/plastoquinone (complex I), various chains
This family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane. Number of members: 1824

[1]
Medline: 93110040
The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. Walker JE; Q Rev Biophys 1992;25:253-324.

387. (oxidored q3) NADH-ubiquinone/plastoquinone oxidoreductase chain 6. 179 members.

388. (oxidored q5) NADH-ubiquinone oxidoreductase chain 4, amino terminus

[1] Walker JE ; Q Rev Biophys 1992;25:253-324.

389. (oxidored q6) Respiratory-chain NADH dehydrogenase 20 Kd subunit signature
Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 20 Kd (in mammals) [3], which is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 20 Kd subunit has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in Neurospora crassa. - Mitochondrial encoded in Paramecium (gene psbG).
- Chloroplast encoded in various higher plants (gene ndhK or psbG).

The 20 Kd subunit is highly similar to [4]:

- Synechocystis strain PCC 6803 proteins psbG1 and psbG2.
- Subunit B of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoB).
- Subunit NQO6 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
- Subunit 7 of Escherichia coli formate hydrogenlyase (gene hycG).
- Subunit I of Escherichia coli hydrogenase-4 (gene hyfI).

As as signature pattern a highly conserved region was selected, located in the central section of this subunit and which contains a conserved cysteine that is probably involved in the binding of the 4Fe-4S center.

Consensus pattern: [GN]-x-D-[KRST]-[LIVMF](2)-P-[IV]-D-[LIVMFYW](2)-x-P-x-C-P-[PT] [The C is a putative 4Fe-4S ligand]

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
[3] Arizmendi J.M., Runswick M.J., Skehel J.M., Walker J.E. FEBS Lett. 301:237-242(1992).
[4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

390. p53 tumor antigen signature

The p53 tumor antigen [1 to 5, E1,E2] is a protein found in increased amounts in a wide variety of transformed cells. It is also detectable in many proliferating nontransformed cells, but it is undetectable or present at low levels in resting cells. It is frequently mutated or inactivated in many types of cancer. p53 seems to act as a tumor suppressor in some, but probably not all, tumor types. p53 is probably involved in cell cycle regulation, and may be a trans-activator that acts to negatively regulate cellular division by controlling a set of genes required for this process.

p53 is a phosphoprotein of about 390 amino acids which can be subdivided into four domains: a highly charged acidic region of about 75 to 80 residues, a hydrophobic proline-rich domain (position 80 to 150), a central region (from 150 to about 300), and a highly basic C-terminal region. The sequence of p53 is well conserved in vertebrate species; attempts to identify p53 in other eukaryotic philum has so far been unsuccessful.

As a signature pattern for p53 a perfectly conserved stretch of 13 residues located in the central region of the protein was selected. This region, known as domain IV in [3], is involved (along with an adjacent region) in the binding of the large T antigen of SV40. In man this region is the focus of a variety of point mutations in cancerous tumors.

Consensus pattern: M-C-N-S-S-C-M-G-G-M-N-R-R

[1] Levine A.J., Momand J., Finlay C.A. Nature 351:453-456(1991).
[2] Levine A.J., Momand J. Biochim. Biophys. Acta 1032:119-136(1990).
[3] Soussi T., Caron De Fromentel C., May P. Oncogene 5:945-952(1990).
[4] Lane D.P., Benchimol S. Genes Dev. 4:1-8(1990).
[5] Ulrich S.J., Anderson C.W., Mercer W.E., Appella E. J. Biol. Chem. 267:15259-15262(1992).

391. (P5CR) Delta 1-pyrroline-5-carboxylate reductase signature

Delta 1-pyrroline-5-carboxylate reductase (P5CR) (EC 1.5.1.2) [1,2] is the enzyme that catalyzes the terminal step in the biosynthesis of proline from glutamate, the NAD(P) dependent oxidation of 1-pyrroline-5-carboxylate into proline.

The sequences of P5CR from eubacteria (gene proC), archaebacteria and eukaryotes show only a moderate level of overall similarity. As a signature pattern, the best conserved region located in the C-terminal section of P5CR was selected.

Consensus pattern: [PALF]-x(2,3)-[LIV]-x(3)-[LIVM]-[STAC]-[STV]-x-[GAN]-G-x-T-x(2)-[AG]-[LIV]-x(2)-[LMF]-[DENQK]

[1] Delauney A.J., Verma D.P. Mol. Gen. Genet. 221:299-305(1990).
[2] Savioz A., Jeenes D.J., Kocher H.P., Haas D. Gene 86:107-111(1990).

392. Poly-adenylate binding protein, unique domain.

393. (PAL) Phenylalanine and histidine ammonia-lyases active site

Phenylalanine ammonia-lyase (EC 4.3.1.5) (PAL) is a key enzyme of plant and fungi phenylpropanoid metabolism which is involved in the biosynthesis of a wide variety of secondary metabolites such as flavanoids, furanocoumarin phytoalexins and cell wall components. These compounds have many important roles in plants during normal growth and in responses to environmental stress. PAL catalyzes the removal of an ammonia group from phenylalanine to form trans-cinnamate.

Histidine ammonia-lyase (EC 4.3.1.3) (histidase) catalyzes the first step in histidine degradation, the removal of an ammonia group from histidine to produce urocanic acid.

The two types of enzymes are functionally and structurally related [1]. They are the only enzymes which are known to have the modified amino acid dehydro-alanine (DHA) in their active site. A serine residue has been shown [2,3,4] to be the precursor of this essential electrophilic moiety. The region around this active site residue is well conserved and can be used as a signature pattern.

Consensus pattern: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SA] [S is the active site residue]

[1] Taylor R.G., Lambert M.A., Sexsmith E., Sadler S.J., Ray P.N., Mahuran D.J., McInnes R.R. J. Biol. Chem. 265:18192-18199(1990).

[2] Langer M., Reck G., Reed J., Retey J. Biochemistry 33:6462-6467(1994).

[3] Schuster B., Retey J. FEBS Lett. 349:252-254(1994).

[4] Taylor R.G., McInnes R.R. J. Biol. Chem. 269:27473-27477(1994).

394. PAS domain

-!- CAUTION. This family does not currently match all known examples of PAS domains.

PAS motifs appear in archaea, eubacteria and eukarya. Probably the most surprising identification of a PAS domain was that in

EAG-like K+-channels[1,3].

Number of members: 308

[1]
Medline: 97446881
PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox.
Zhulin IB, Taylor BL, Dixon R;
Trends Biochem Sci 1997;22:331-333.
[2]Medline: 95275818
1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore.
Borgstahl GE, Williams DR, Getzoff ED;
Biochemistry 1995;34:6278-6287.
[3]Medline: 98044337
PAS: a multifunctional domain family comes to light.
Ponting CP, Aravind L;
Curr Biol 1997;7:674-677.

395. (PBP) Phosphatidylethanolamine-binding protein family signature

Mammalian phosphatidylethanolamine-binding protein (also knowns as basic cytosolic 21 Kd protein) is a 186 residue protein found in a variety of tissues [1]. It binds hydrophobic ligands, such as phosphatidylethanolamine, but also seems [2] to bind nucleotides such as GTP and FMN, it is suggested that it could act in membrane remodeling during growth and maturation. This protein belongs to a family that also includes:

- Drosophila antennal protein A5, a putative odorant-binding protein.
- Onchocerca volvulus antigen Ov-16 and the related proteins D1, D2 and D3.
- Plasmodium falciparum putative phosphatidylethanolamine-binding protein.
- Toxocara canis secreted antigen TES-26. This larval protein has been shown to bind phosphatidylethanolamine.
- Yeast protein DKA1 (also known as NSP1 or TFS1). The function of this protein is not very clear. - Yeast hypothetical protein YLR179C.
- Caenorhabditis elegans hypothetical protein F40A3.3.

As a signature pattern, the best conserved region was selected which is located in the end of the first third of the sequence of these proteins.

Consensus pattern: [FYL]-x-[LV]-[LIVF]-x-[TIV]-[DC]-P-D-x-P-[SN]-x(10)-H

[1] Seddiqi N., Bollengier F., Alliel P.M., Perin J.P., Bonnet F., Bucquoy S., Jolles P., Schoentgen F. J. Mol. Evol. 39:655-660(1994).

[2] Schoentgen F., Jolles P. FEBS Lett. 369:22-6(1995).

396. PCI domain

This domain has also been called the PINT motif (Proteasome, Int-6, Nip-1 and TRIP-15) [1].

Number of members: 49

[1]

Medline: 98308842

The PCI domain: a common theme in three multiprotein complexes.

Hofmann K, Bucher P;

Trends Biochem Sci 1998;23:204-205.

[2]Medline: 98266368

Homologues of 26S proteasome subunits are regulators of transcription and translation.

Aravind L, Ponting CP;

Protein Sci 1998;7:1250-1254.

397. (PCMT) Protein-L-isoaspartate (D-aspartate) O-methyltransferase signature. Protein-L-isoaspartate (D-aspartate) O-methyltransferase (EC 2.1.1.77) (PCMT)[1] (which is also known as L-isoaspartyl protein carboxyl methyltransferase)is an enzyme that catalyzes the transfer of a methyl group from S-adenosylmethionine to the free carboxyl groups of D-aspartyl or L-isoaspartyl residues in a variety of peptides and proteins. The enzyme does not act on normal L-aspartyl residues L-isoaspartyl and D-aspartyl are the products of the spontaneous de amidation and/or isomerization of normal L-aspartyl and L-asparaginyl

residues in proteins. PCMT plays a role in the repair and/ordegradation of these damaged proteins; the enzymatic methyl esterification of the abnormal residues can lead to their conversion to normal L-aspartylresidues. PCMT is a well-conserved and widely distributed cytosolic protein of about 24Kd. As a signature pattern, a conserved region in the central part of this enzyme has been developed.

Consensus pattern: [GSA]-D-G-x(2)-G-[FYWV]-x(3)-[AS]-P-[FY]-[DN]-x-I –

[1] Kagan R.M., McFadden H.J., McFadden P.N., O'Connor C., Clarke S. Comp. Biochem. Physiol. 117b:379-385(1997).

398. (PCNA) Proliferating cell nuclear antigen signatures

Proliferating cell nuclear antigen (PCNA) [1,2] is a protein involved in DNA replication by acting as a cofactor for DNA polymerase delta, the polymerase responsible for leading strand DNA replication.

A similar protein exists in yeast (gene POL30) [3] and is associated with polymerase III, the yeast analog of polymerase delta. In baculoviruses the ETL protein has been shown [4] to be highly related to PCNA and is probably associated with the viral encoded DNA polymerase. An homolog of PCNA is also found in archebacteria.

As signatures for this family of proteins, two conserved regions were selected located in the N-terminal section. The second one has been proposed to bind DNA.

Consensus pattern: [GA]-[LIVMF]-x-[LIVMA]-x-[SAV]-[LIVM]-D-x-[NSAE]-[HKR]-[VI]-x-[LY]-[VGA]-x-[LIVM]-x-[LIVM]-x(4)-F

-Consensus pattern: [RKA]-C-[DE]-[RH]-x(3)-[LIVMF]-x(3)-[LIVM]-x-[SGAN]-[LIVMF]-x-K-[LIVMF](2)

[1] Bravo R., Frank R., Blundell P.A., McDonald-Bravo H. Nature 326:515-517(1987).

[2] Suzuka I., Hata S., Matsuoka M., Kosugi S., Hashimoto J. Eur. J. Biochem. 195:571-575(1991).[3] Bauer G.A., Burgess P.M.J. Nucleic Acids Res. 18:261-265(1990).

[4] O'Reilly D.R., Crawford A.M., Miller L.K. Nature 337:606-606(1989).

399. (PDT) Prephenate dehydratase signatures

Prephenate dehydratase (EC 4.2.1.51) (PDT) catalyzes the decarboxylation of prephenate into phenylpyruvate. In microorganisms PDT is involved in the terminal pathway of the biosynthesis of phenylalanine. In some bacteria such as Escherichia coli PDT is part of a bifunctional enzyme (P-protein) that also catalyzes the transformation of chorismate into prephenate (chorismate mutase) while in other bacteria it is a monofunctional enzyme. The sequence of monofunctional PDT align well with the C-terminal part of that of P-proteins [1].

As signature patterns for PDT two conserved regions were selected. The first region contains a conserved threonine which has been said to be essential for the activity of the enzyme in E. coli. The second region includes a conserved glutamate. Both regions are in the C-terminal part of PDT.

Consensus pattern: [FY]-x-[LIVM]-x(2)-[LIVM]-x(5)-[DN]-x(5)-T-R-F-[LIVMW]-x-[LIVM]

[1] Fischer R.S., Zhao G., Jensen R.A. J. Gen. Microbiol. 137:1293-1301(1991).

400. PDZ domain (Also known as DHR or GLGF).

PDZ domains are found in diverse signaling proteins.

[1] Ponting CP, Phillips C, Davies KE, Blake DJ Bioessays 1997;19:469-479. [2] Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R; Cell. 1996;85:1067-1076. [3] Ponting CP; Protein Sci 1997;6:464-468.

401. (PPDK_N_term) PEP-utilizing enzymes signatures

A number of enzymes that catalyze the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) via a phospho-histidine intermediate have been shown to be structurally related [1,2,3,4]. These enzymes are:

- Pyruvate,orthophosphate dikinase (EC 2.7.9.1) (PPDK). PPDK catalyzes the

reversible phosphorylation of pyruvate and phosphate by ATP to PEP and diphosphate. In plants PPDK function in the direction of the formation of PEP, which is the primary acceptor of carbon dioxide in C4 and crassulacean acid metabolism plants. In some bacteria, such as Bacteroides symbiosus, PPDK functions in the direction of ATP synthesis.

- Phosphoenolpyruvate synthase (EC 2.7.9.2) (pyruvate,water dikinase). This enzyme catalyzes the reversible phosphorylation of pyruvate by ATP to form PEP, AMP and phosphate, an essential step in gluconeogenesis when pyruvate and lactate are used as a carbon source.
- Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9). This is the first enzyme of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS), a major carbohydrate transport system in bacteria. The PTS catalyzes the phosphorylation of incoming sugar substrates concomitant with their translocation across the cell membrane. The general mechanism of the PTS is the following: a phosphoryl group from PEP is transferred to enzyme-I (EI) of PTS which in turn transfers it to a phosphoryl carrier protein (HPr). Phospho-HPr then transfers the phosphoryl group to a sugar-specific permease.

All these enzymes share the same catalytic mechanism: they bind PEP and transfer the phosphoryl group from it to a histidine residue. The sequence around that residue is highly conserved and can be used as a signature pattern for these enzymes. As a second signature pattern a conserved region was selected in the C-terminal part of the PEP-utilizing enzymes. The biological significance of this region is not yet known.

Consensus pattern: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG] [H is phosphorylated]

-Consensus pattern: [DEQSK]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-[LIVMFYGT]-[STALIV]-[LIVMF]-[GAS]-x(2)-R

[1] Reizer J., Hoischen C., Reizer A., Pham T.N., Saier M.H. Jr. Protein Sci. 2:506-521(1993).

[2] Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J., Saier M.H. Jr. Gene 181:103-108(1996).

[3] Pocalyko D.J., Carroll L.J., Martin B.M., Babbitt P.C., Dunaway-Mariano D. Biochemistry 29:10757-10765(1990).

[4] Niersbach M., Kreuzaler F., Geerse R.H., Postma P., Hirsch H.J. Mol. Gen. Genet. 232:332-336(1992).

402. (PEPCK ATP) Phosphoenolpyruvate carboxykinase (ATP) signature Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) (PEPCK) [1] catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate while hydrolyzing ATP, a rate limiting step in gluconeogenesis (the biosynthesis of glucose).

The sequence of this enzyme has been obtained from Escherichia coli, yeast, and Trypanosoma brucei; these three sequences are evolutionary related and share many regions of similarity. As a signature pattern a highly conserved region was selected that contains four acidic residues and which is located in the central part of the enzyme. The beginning of the pattern is located about 10 residues to the C-terminus of an ATP-binding motif 'A' (P-loop) (see <PDOC00017>) and is also part of the ATP-binding domain [2].

Consensus pattern: L-I-G-D-D-E-H-x-W-x-[DE]-x-G-[IV]-x-N

-Note: phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) an enzyme that catalyzes the same reaction, but using GTP instead of ATP, is not related to the above enzyme (see <PDOC00421>).

[1] Medina V., Pontarollo R., Glaeske D., Tabel H., Goldie H. J. Bacteriol. 172:7151-7156(1990).

[2] Matte A., Goldie H., Sweet R.M., Delbaere L.T.J. J. Mol. Biol. 256:126-143(1996).

403. (Pepcase) Phosphoenolpyruvate carboxylase active sites. Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPcase) catalyzes the irreversible beta-carboxylation of phosphoenolpyruvate by bicarbonate to yield oxaloacetate and phosphate. The enzyme is found in all plants and in a variety of microorganisms. A histidine [1] and a lysine [2] have been implicated in the catalytic mechanism of this enzyme; the regions around these active

site residues are highly conserved in PEPcase from various plants, bacteria and cyanobacteria and can be used as a signature patterns for this type of enzyme.

Consensus pattern: [VT]-x-T-A-H-P-T-[EQ]-x(2)-R-[KRH] [H is an active site residue]-
Consensus pattern: [IV]-M-[LIVM]-G-Y-S-D-S-x-K-D-[STAG]-G [K is an active site residue]-

[1] Terada K., Izui K. Eur. J. Biochem. 202:797-803(1991).[2] Jiao J.-A., Podesta F.E., Chollet R., O'Leary M.H., Andreo C.S. Biochim. Biophys. Acta 1041:291-295(1990).

404. PET112 family signature

The following proteins from eukaryotes, prokaryotes and archaebacteria belong to the same family:

- Yeast mitochondrial protein PET112 [1], which plays an unknown role in the expression of mitochondrial genes, probably at the level of translation.
- Aspergillus nidulans mitochondrial protein nempA.
- Bacillus subtilis hypothetical protein yzdD.
- Moraxella catarrhalis hypothetical protein in bloR-1 3'region.
- Mycoplasma genitalium hypothetical protein MG100.
- Methanococcus jannaschii hypothetical proteins MJ0019 and MJ0160.

The size of these proteins range from 419 to 630 amino acids. As a signature pattern, a conserved region located in the N-terminal section was selected.

Consensus pattern: [DN]-x-[DN]-R-x(3)-P-L-[LIV]-E-[LIV]-x-[ST]-x-P

[1] Mulero J.J., Rosenthal J.K., Fox T.D. Curr. Genet. 25:299-304(1994).

405. (PFK) Phosphofructokinase signature

Phosphofructokinase (EC 2.7.1.11) (PFK) [1,2] is a key regulatory enzyme in the glycolytic pathway. It catalyzes the phosphorylation by ATP of fructose 6-phosphate to fructose 1,6-bisphosphate. In bacteria PFK is a tetramer of identical 36 Kd subunits. In mammals it is a tetramer of 80 Kd subunits. Each

80 Kd subunit consist of two homologous domains which are highly related to the bacterial 36 Kd subunits. In Human there are three, tissue-specific, types of PFK isozymes: PFKM (muscle), PFKL (liver), and PFKP (platelet). In yeast PFK is an octamer composed of four 100 Kd alpha chains (gene PFK1) and four 100 Kd beta chains (gene PFK2); like the mammalian 80 Kd subunits, the yeast 100 Kd subunits are composed of two homologous domains.
As a signature pattern for PFK a region that contains three basic residues involved in fructose-6-phosphate binding was selected.

Consensus pattern: [RK]-x(4)-G-H-x-Q-[QR]-G-G-x(5)-D-R [The R/K, the H and the Q/R are involved in fructose-6-P binding]
-Note: Escherichia coli has two phosphofructokinase isozymes which are encoded by genes pfkA (major) and pfkB (minor). The pfkB isozyme is not evolutionary related to other prokaryotic or eukaryotic PFK's (see <PDOC00504>).

[1] Poorman R.A., Randolph A., Kemp R.G., Heinrikson R.L. Nature 309:467-469(1984).
[2] Heinisch J., Ritzel R.G., von Borstel R.C., Aguilera A., Rodicio R., Zimmermann F.K. Gene 78:309-321(1989).

406. (PGAM) Phosphoglycerate mutase family phosphohistidine signature
Phosphoglycerate mutase (EC 5.4.2.1) (PGAM) and bisphosphoglycerate mutase (EC 5.4.2.4) (BPGM) are structurally related enzymes which catalyze reactions involving the transfer of phospho groups between the three carbon atoms of phosphoglycerate [1,2]. Both enzymes can catalyze three different reactions, although in different proportions:
- The isomerization of 2-phosphoglycerate (2-PGA) to 3-phosphoglycerate (3-PGA) with 2,3-diphosphoglycerate (2,3-DPG) as the primer of the reaction.
- The synthesis of 2,3-DPG from 1,3-DPG with 3-PGA as a primer.
- The degradation of 2,3-DPG to 3-PGA (phosphatase EC 3.1.3.13 activity).

In mammals, PGAM is a dimeric protein. There are two isoforms of PGAM: the M (muscle) and B (brain) forms. In yeast, PGAM is a tetrameric protein. BPGM is a dimeric protein and is found mainly in erythrocytes where it plays a major role in regulating hemoglobin oxygen affinity as a consequence of controlling

2,3-DPG concentration.

The catalytic mechanism of both PGAM and BPGM involves the formation of a phosphohistidine intermediate [3].

The bifunctional enzyme 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase (EC 2.7.1.105 and EC 3.1.3.46) (PF2K) [4] catalyzes both the synthesis and the degradation of fructose-2,6-bisphosphate. PF2K is an important enzyme in the regulation of hepatic carbohydrate metabolism. Like PGAM/BPGM, the fructose-2,6-bisphosphatase reaction involves a phosphohistidine intermediate and the phosphatase domain of PF2K is structurally related to PGAM/BPGM.

The bacterial enzyme alpha-ribazole-5'-phosphate phosphatase (gene cobC) which is involved in cobalamin biosynthesis also belongs to this family [5].

A signature pattern was built around the phosphohistidine residue.

Consensus pattern: [LIVM]-x-R-H-G-[EQ]-x(3)-N [H is the phosphohistidine residue]

-Note: some organisms harbor a form of PGAM independent of 2,3-DPG, this enzyme is not related to the family described above [6].

[1] Le Boulch P., Joulin V., Garel M.-C., Rosa J., Cohen-Solal M. Biochem. Biophys. Res. Commun. 156:874-881(1988).

[2] White M.F., Fothergill-Gilmore L.A. FEBS Lett. 229:383-387(1988).

[3] Rose Z.B. Meth. Enzymol. 87:43-51(1982).

[4] Bazan J.F., Fletterick R.J., Pilkis S.J. Proc. Natl. Acad. Sci. U.S.A. 86:9642-9646(1989).

[5] O'Toole G.A., Trzebiatowski J.R., Escalante-Semerena J.C. J. Biol. Chem. 269:26503-26511(1994).

[6] Grana X., De Lecea L., El-Maghrabi M.R., Urena J.M., Caellas C., Carreras J., Puigdomenech P., Pilkis S.J., Climent F. J. Biol. Chem. 267:12797-12803(1992).

407. (PGI) Phosphoglucose isomerase signatures

Phosphoglucose isomerase (EC 5.3.1.9) (PGI) [1,2] is a dimeric enzyme that catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate. PGI is involved in different pathways: in most higher organisms it is involved in glycolysis; in mammals it is involved in gluconeogenesis; in

plants in carbohydrate biosynthesis; in some bacteria it provides a gateway for fructose into the Entner-Doudouroff pathway. PGI has been shown [3] to be identical to neuroleukin, a neurotrophic factor which supports the survival of various types of neurons.

The sequence of PGI from many species ranging from bacteria to mammals is available and has been shown to be highly conserved. As signature patterns for this enzyme two conserved regions were selected, the first region is located in the central section of PGI, while the second one is located in its C-terminal section.

Consensus pattern: [DENS]-x-[LIVM]-G-G-R-[FY]-S-[LIVMT]-x-[STA]-[PSAC]-[LIVMA]-G

-Consensus pattern: [GS]-x-[LIVM]-[LIVMFYW]-x(4)-[FY]-[DN]-Q-x-G-V-E-x(2)-K

[1] Achari A., Marshall S.E., Muirhewad H., Palmieri R.H., Noltmann E.A. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 293:145-157(1981).

[2] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:544-545(1992).

[3] Faik P., Walker J.I.H., Redmill A.A.M., Morgan M.J. Nature 332:455-456(1988).

408. (PGK) Phosphoglycerate kinase signature

Phosphoglycerate kinase (EC 2.7.2.3) (PGK) [1] catalyzes the second step in the second phase of glycolysis, the reversible conversion of 1,3-diphospho-glycerate to 3-phosphoglycerate with generation of one molecule of ATP. PGK is found in all living organisms and its sequence has been highly conserved throughout evolution. It is a two-domain protein; each domain is composed of six repeats of an alpha/beta structural motif. As a signature pattern for PGK's, a conserved region in the N-terminal region was selected.

Consensus pattern: [KRHGTCVN]-[VT]-[LIVMF]-[LIVMC]-R-x-D-x-N-[SACV]-P

[1] Watson H.C., Littlechild J.A. Biochem. Soc. Trans. 18:187-190(1990).

409. (PGM PMM) Phosphoglucomutase and phosphomannomutase phosphoserine signature

- Phosphoglucomutase (EC 5.4.2.2) (PGM). PGM is an enzyme responsible for the conversion of D-glucose 1-phosphate into D-glucose 6-phosphate. PGM participates in both the breakdown and synthesis of glucose [1].
- Phosphomannomutase (EC 5.4.2.8) (PMM). PMM is an enzyme responsible for the conversion of D-mannose 1-phosphate into D-mannose 6-phosphate. PMM is required for different biosynthetic pathways in bacteria. For example, in enterobacteria such as Escherichia coli there are two different genes coding for this enzyme: rfbK which is involved in the synthesis of the O antigen of lipopolysaccharide and cpsG which is required for the synthesis of the M antigen capsular polysaccharide [2]. In Pseudomonas aeruginosa PMM (gene algC) is involved in the biosynthesis of the alginate layer [3] and in Xanthomonas campestris (gene xanA) it is involved in the biosynthesis of xanthan [4]. In Rhizobium strain ngr234 (gene noeK) it is involved in the biosynthesis of the nod factor.
- Phosphoacetylglucosamine mutase (EC 5.4.2.3) which converts N-acetyl-D-glucosamine 1-phosphate into the 6-phosphate isomer.

The catalytic mechanism of both PGM and PMM involves the formation of a phosphoserine intermediate [1]. The sequence around the serine residue is well conserved and can be used as a signature pattern.

In addition to PGM and PMM there are at least three uncharacterized proteins that belong to this family [5,6]:

- Urease operon protein ureC from Helicobacter pylori.
- Escherichia coli protein mrsA.
- Paramecium tetraurelia parafusin, a phosphoglycoprotein involved in exocytosis.
- A Methanococcus vannielii hypothetical protein in the 3'region of the gene for ribosomal protein S10.

Consensus pattern: [GSA]-[LIVM]-x-[LIVM]-[ST]-[PGA]-S-H-x-P-x(4)-[GNHE] [S is the phosphoserine residue]

-Note: PMM from fungi do not belong to this family.

[1] Dai J.B., Liu Y., Ray W.J. Jr., Konno M. J. Biol. Chem. 267:6322-6337(1992).

[2] Stevenson G., Lee S.J., Romana L.K., Reeves P.R. Mol. Gen. Genet. 227:173-180(1991).

[3] Zielinski N.A., Chakrabarty A.M., Berry A. J. Biol. Chem. 266:9754-9763(1991).

[4] Koeplin R., Arnold W., Hoette B., Simon R., Wang G., Puehler A. J. Bacteriol. 174:191-199(1992).

[5] Bairoch A. Unpublished observations (1993).

[6] Subramanian S.V., Wyroba E., Andersen A.P., Satir B.H. Proc. Natl. Acad. Sci. U.S.A. 91:9832-9836(1994).

410. PH domain profile

The 'pleckstrin homology' (PH) domain is a domain of about 100 residues that occurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton [1 to 7].

The function of this domain is not clear, several putative functions have been suggested: - binding to the beta/gamma subunit of heterotrimeric G proteins,

- binding to lipids, e.g. phosphatidylinositol-4,5-bisphosphate,
- binding to phosphorylated Ser/Thr residues,
- attachment to membranes by an unknown mechanism.

It is possible that different PH domains have totally different ligand requirements.

The 3D structure of several PH domains has been determined [8]. All known cases have a common structure consisting of two perpendicular anti-parallel beta sheets, followed by a C-terminal amphipathic helix. The loops connecting the beta-strands differ greatly in length, making the PH domain relatively difficult to detect. There are no totally invariant residues within the PH domain.

Proteins reported to contain one more PH domains belong to the following families:

- Pleckstrin, the protein where this domain was first detected, is the major substrate of protein kinase C in platelets. Pleckstrin is one of the rare proteins to contains two PH domains.
- Ser/Thr protein kinases such as the Act/Rac family, the beta-adrenergic receptor kinases, the mu isoform of PKC and the trypanosomal NrkA family.

- Tyrosine protein kinases belonging to the Btk/ltk/Tec subfamily.
- Insulin Receptor Substrate 1 (IRS-1).
- Regulators of small G-proteins like guanine nucleotide releasing factor GNRP (Ras-GRF) (which contains 2 PH domains), guanine nucleotide exchange proteins like vav, dbl, SoS and yeast CDC24, GTPase activating proteins like rasGAP and BEM2/IPL2, and the human break point cluster protein bcr.
- Cytoskeletal proteins such as dynamin (see <PDOC00362>), Caenorhabditis elegans kinesin-like protein unc-104 (see <PDOC00343>), spectrin beta-chain, syntrophin (2 PH domains) and yeast nuclear migration protein NUM1.
- Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) (see <PDOC50007>) isoforms gamma and delta. Isoform gamma contains two PH domains, the second one is split into two parts separated by about 400 residues. - Oxysterol binding proteins OSBP, yeast OSH1 and YHR073w.
- Mouse protein citron, a putative rho/rac effector that binds to the GTP-bound forms of rho and rac,
- Several yeast proteins involved in cell cycle regulation and bud formation like BEM2, BEM3, BUD4 and the BEM1-binding proteins BOI2 (BEB1) and BOI1 (BOB1). - Caenorhabditis elegans protein MIG-10.
- Caenorhabditis elegans hypothetical proteins C04D8.1, K06H7.4 and ZK632.12.
- Yeast hypothetical proteins YBR129c and YHR155w.

The profile for the PH domain, which has been developed by Toby Gibson at the EMBL, covers the total length of domain. Several proteins contain large insertions in the PH domain and are thus difficult to detect with this profile. In some of these cases, the profile will align only to one half of the PH domain.

-Sequences known to belong to this class detected by the pattern: ALL. But it should be noted that while all sequences containing PH domains are detected, not all PH domains are. Some of the split domains lie below the cutoff threshold.

[1] Mayer B.J., Ren R., Clark K.L., Baltimore D. Cell 73:629-630(1993).
[2] Haslam R.J., Koide H.B., Hemmings B.A. Nature 363:309-310(1993).
[3] Musacchio A., Gibson T.J., Rice P., Thompson J., Saraste M. Trends Biochem. Sci. 18:343-348(1993).
[4] Gibson T.J., Hyvonen M., Musacchio A., Saraste M., Birney E.

Trends Biochem. Sci. 19:349-353(1994).[5] Pawson T. Nature 373:573-580(1995).[6] Ingley E., Hemmings B.A. J. Cell. Biochem. 56:436-443(1994).[7] Saraste M., Hyvonen M. Curr. Opin. Struct. Biol. 5:403-408(1995).[8] Riddihough G. Nat. Struct. Biol. 1:755-757(1994).

411. PHD-finger

[1]

Medline: 95216093

The PHD finger: implications for chromatin-mediated transcriptional regulation.

Aasland R, Gibson TJ, Stewart AF;

Trends Biochem Sci 1995;20:56-59.

Number of members: 181

412. (PI-PLC-X) Phosphatidylinositol-specific phospholipase C profiles

Phosphatidylinositol-specific phospholipase C (EC 3.1.4.11), an eukaryotic intracellular enzyme, plays an important role in signal transduction processes [1]. It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5-triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins [2 to 4].

In mammals, there are at least 6 different isoforms of PI-PLC, they differ in their domain structure, their regulation, and their tissue distribution. Lower eukaryotes also possess multiple isoforms of PI-PLC.

All eukaryotic PI-PLCs contain two regions of homology, sometimes referred to as 'X-box' and 'Y-box'. The order of these two regions is always the same (NH2-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is

a C2 domain (see <PDOC00380>) possibly involved in Ca-dependent membrane attachment.

Profile analysis shows that sequences with significant similarity to the X-box domain occur also in prokaryotic and trypanosome PI-specific phospholipases C. Apart from this region, the prokaryotic enzymes show no similarity to their eukaryotic counterparts.

Two profiles were developed, one covering the X-box, the other the Y-box.

[1] Meldrum E., Parker P.J., Carozzi A.
Biochim. Biophys. Acta 1092:49-71(1991).[2] Rhee S.G., Choi K.D.
Adv. Second Messenger Phosphoprotein Res. 26:35-61(1992).
[3] Rhee S.G., Choi K.D. J. Biol. Chem. 267:12393-12396(1992).
[4] Sternweis P.C., Smrcka A.V. Trends Biochem. Sci. 17:502-506(1992).

413. (PI-PLC-Y) Phosphatidylinositol-specific phospholipase C profiles

Phosphatidylinositol-specific phospholipase C (EC 3.1.4.11), an eukaryotic intracellular enzyme, plays an important role in signal transduction processes [1]. It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5-triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins [2 to 4].

In mammals, there are at least 6 different isoforms of PI-PLC, they differ in their domain structure, their regulation, and their tissue distribution. Lower eukaryotes also possess multiple isoforms of PI-PLC.

All eukaryotic PI-PLCs contain two regions of homology, sometimes referred to as 'X-box' and 'Y-box'. The order of these two regions is always the same (NH2-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain (see <PDOC00380>) possibly involved in Ca-dependent membrane attachment.

Profile analysis shows that sequences with significant similarity

to the X-box domain occur also in prokaryotic and trypanosome PI-specific phospholipases C. Apart from this region, the prokaryotic enzymes show no similarity to their eukaryotic counterparts.

Two profiles were developed, one covering the X-box, the other the Y-box.

[1] Meldrum E., Parker P.J., Carozzi A.
Biochim. Biophys. Acta 1092:49-71(1991).[2] Rhee S.G., Choi K.D.
Adv. Second Messenger Phosphoprotein Res. 26:35-61(1992).
[3] Rhee S.G., Choi K.D. J. Biol. Chem. 267:12393-12396(1992).
[4] Sternweis P.C., Smrcka A.V. Trends Biochem. Sci. 17:502-506(1992).

414. (PK) Pyruvate kinase active site signature

Pyruvate kinase (EC 2.7.1.40) (PK) [1] catalyzes the final step in glycolysis, the conversion of phosphoenolpyruvate to pyruvate with the concomitant phosphorylation of ADP to ATP. PK requires both magnesium and potassium ions for its activity. PK is found in all living organisms. In vertebrates there are four, tissues specific, isozymes: L (liver), R (red cells), M1 (muscle, heart, and brain), and M2 (early fetal tissues). In Escherichia coli there are two isozymes: PK-I (gene pykF) and PK-II (gene pykA). All PK isozymes seem to be tetramers of identical subunits of about 500 amino acid residues.

As a signature pattern for PK a conserved region was selected that includes a lysine residue which seems to be the acid/base catalyst responsible for the interconversion of pyruvate and enolpyruvate, and a glutamic acid residue implicated in the binding of the magnesium ion.

Consensus pattern: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQHS]-[GSTA]-[LIVM] [K is the active site residue] [E is a magnesium ligand]

[1] Muirhead H. Biochem. Soc. Trans. 18:193-196(1990).

415. (PLDc) Phospholipase D. Active site motif

Phosphatidylcholine-hydrolyzing phospholipase D (PLD) isoforms are activated by ADP-ribosylation factors (ARFs). PLD produces phosphatidic

acid from phosphatidylcholine, which may be essential for the formation of certain types of transport vesicles or may be constitutive vesicular transport to signal transduction pathways.

PC-hydrolyzing PLD is a homologue of cardiolipin synthase, phosphatidylserine synthase, bacterial PLDs, and viral proteins. Each of these appears to possess a domain duplication which is apparent by the presence of two motifs containing well-conserved histidine, lysine, and/or asparagine residues which may contribute to the active site. aspartic acid. An E. coli endonuclease (nuc) and similar proteins appear to be PLD homologues but possess only one of these motifs. The profile contained here represents only the putative active site regions, since an accurate multiple alignment of the repeat units has not been achieved.

Number of members: 139

[1]
Medline: 96303814
A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues.
Ponting CP, Kerr ID;
Protein Sci 1996;5:914-922.

[2]Medline: 96334293
A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins.
Koonin EV;
Trends Biochem Sci 1996;21:242-243.

[3]Medline: 94327597
Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from Ricinus communis L.
Wang X, Xu L, Zheng L;
J Biol Chem 1994;269:20312-20317.

[4]Medline: 97386825

Regulation of eukaryotic phosphatidylinositol-specific phospholipase C and phospholipase D.
Singer WD, Brown HA, Sternweis PC;
Annu Rev Biochem 1997;66:475-509.

416. (PMI type1) Phosphomannose isomerase type I signatures

Phosphomannose isomerase (EC 5.3.1.8) (PMI) [1,2] is the enzyme that catalyzes the interconversion of mannose-6-phosphate and fructose-6-phosphate. In eukaryotes, it is involved in the synthesis of GDP-mannose which is a constituent of N- and O-linked glycans as well as GPI anchors. In prokaryotes, it is involved in a variety of pathways including capsular polysaccharide biosynthesis and D-mannose metabolism.

Three classes of PMI have been defined on the basis of sequence similarities [1]. The first class comprises all known eukaryotic PMI as well as the enzyme encoded by the manA gene in enterobacteria such as Escherichia coli. Class I PMI's are proteins of about 42 to 50 Kd which bind a zinc ion essential for their activity.

As signature patterns for class I PMI, two conserved regions were selected. The first one is located in the N-terminal section of these proteins, the second in the C-terminal half. Both patterns contain a residue involved [3] in the binding of the zinc ion.

Consensus pattern: Y-x-D-x-N-H-K-P-E [E is a zinc ligand]

-Consensus pattern: H-A-Y-[LIVM]-x-G-x(2)-[LIVM]-E-x-M-A-x-S-D-N-x-[LIVM]-R-A-G-x-T-P-K [H is a zinc ligand]

[1] Proudfoot A.E.I., Turcatti G., Wells T.N.C., Payton M.A., Smith D.J. Eur. J. Biochem. 219:415-423(1994).

[2] Coulin F., Magnenat E., Proudfoot A.E.I., Payton M.A., Scully P., Wells T.N.C. Biochemistry 32:14139-14144(1993).

[3] Cleasby A., Wonacott A., Skarzynski T., Hubbard R.E., Davies G.J., Proudfoot A.E.I., Bernard A.R., Payton M.A., Wells T.N.C. Nat. Struct. Biol. 3:470-479(1996).

417. (PNP UDP 1) Purine and other phosphorylases family 1 signature

The following phosphorylases belongs to the same family:

- Purine nucleoside phosphorylase (EC 2.4.2.1) (PNP) from most bacteria (gene deoD). This enzyme catalyzes the cleavage of guanosine or inosine to respective bases and sugar-1-phosphate molecules [1].
- Uridine phosphorylase (EC 2.4.2.3) (UdRPase) from bacteria (gene udp) and mammals. Catalyzes the cleavage of uridine into uracil and ribose-1-phosphate. The products of the reaction are used either as carbon and energy sources or in the rescue of pyrimidine bases for nucleotide synthesis [2].
- 5'-methylthioadenosine phosphorylase (EC 2.4.2.28) (MTA phosphorylase) from Sulfolobus solfataricus [3].

As a signature pattern, a conserved region was selected in the central part of these enzymes.

Consensus pattern: [GST]-x-G-[LIVM]-G-x-[PA]-S-x-[GSTA]-I-x(3)-E-L

-Note: it shoudl be noted that mammalian and some bacterial PNP as well as eukaryotic MTA phosphorylase belong to a different family of phosphorylases (see <PDOC00954>).

[1] Takehara M., Ling F., Izawa S., Inoue Y., Kimura A. Biosci. Biotechnol. Biochem. 59:1987-1990(1995).

[2] Watanabe S.-I., Hino A., Wada K., Eliason J.F., Uchida T. J. Biol. Chem. 270:12191-12196(1995).

[3] Cacciapuoti G., Porcelli M., Bertoldo C., De Rosa M., Zappia V. J. Biol. Chem. 269:24762-24769(1994).

418. (PP2C) Protein phosphatase 2C signature

Protein phosphatase 2C (PP2C) is one of the four major classes of mammalian serine/threonine specific protein phosphatases (EC 3.1.3.16). PP2C [1] is a monomeric enzyme of about 42 Kd which shows broad substrate specificity and is dependent on divalent cations (mainly manganese and magnesium) for its activity. Its exact physiological role is still unclear. Three isozymes are

currently known in mammals: PP2C-alpha, -beta and -gamma. In yeast, there are at least four PP2C homologs: phosphatase PTC1 [2] which has weak tyrosine phosphatase activity in addition to its activity on serines, phosphatases PTC2 and PTC3, and hypothetical protein YBR125c. Isozymes of PP2C are also known from Arabidopsis thaliana (ABI1, PPH1), Caenorhabditis elegans (FEM-2, F42G9.1, T23F11.1), Leishmania chagasi and Paramecium tetraurelia.

In Arabidopsis thaliana, the kinase associated protein phosphatase (KAPP) [3] is an enzyme that dephosphorylates the Ser/Thr receptor-like kinase RLK5 and which contains a C-terminal PP2C domain.

PP2C does not seem to be evolutionary related to the main family of serine/threonine phosphatases: PP1, PP2A and PP2B . However, it is significantly similar to the catalytic subunit of pyruvate dehydrogenase phosphatase (EC 3.1.3.43) (PDPC) [4], which catalyzes dephosphorylation and concomitant reactivation of the alpha subunit of the E1 component of the pyruvate dehydrogenase complex. PDPC is a mitochondrial enzyme and, like PP2C, is magnesium-dependent.

As a signature pattern, the best conserved region was selected which is located in the N-terminal part and contains a perfectly conserved tripeptide. This region includes a conserved aspartate residue involved in divalent cation binding [5].

Consensus pattern: [LIVMFY]-[LIVMFYA]-[GSAC]-[LIVM]-[FYC]-D-G-H-[GAV]

-Note: PP2C belongs [6] to a superfamily which also includes bacterial proteins such as Bacillus spoIIE, rsbU and rsbW, Synechocystis PCC 6803 icfG as well as a domain in fungal adenylate cyclases.

[1] Wenk J., Trompeter H.-I., Pettrich K.-G., Cohen P.T.W., Campbell D.G., Mieskes G. FEBS Lett. 297:135-138(1992).

[2] Maeda T., Tsai A.Y.M., Saito H. Mol. Cell. Biol. 13:5408-5417(1993).

[3] Stone J.M., Collinge M.A., Smith R.D., Horn M.A., Walker J.C. Science 266:793-795(1994).

[4] Lawson J.E., Niu X.-D., Browning K.S., Trong H.L., Yan J., Reed L.J. Biochemistry 32:8987-8993(1993).

[5] Das A.K., Helps N.R., Cohen P.T.W., Barford D. EMBO J. 24:6798-6809(1996).

[6] Bork P., Brown N.P., Hegyi H., Schultz J. Protein Sci. 5:1421-1425(1996).

419. (PPTA) Protein prenyltransferases alpha subunit repeat signature

Protein prenyltransferases catalyze the transfer of an isoprenyl moiety to a cysteine four residues from the C-terminus of several proteins. They are heterodimeric enzymes consisting of alpha and beta subunits. The alpha subunit is thought to participate in a stable complex with the isoprenyl substrate; the beta subunit binds the peptide substrate. Distinct protein prenyltransferases might share a common alpha subunit. Both the alpha and beta subunit show repetitive sequence motifs [1]. These repeats have distinct structural and functional implications and are unrelated to each other. Known protein prenyltransferase alpha subunits are:

- Mammalian protein farnesyltransferase alpha subunit.
- Yeast protein RAM2, a protein farnesyltransferase alpha subunit.
- Yeast protein BET4, a protein geranylgeranyltransferase alpha subunit.

The conserved domain of the alpha subunit consists of about 34 amino acids and is repeated five times. It contains an invariant tryptophan possibly involved in heterodimerization with the conserved phenylalanines in the repeated domains of the beta subunits, via hydrophobic bonds. The signature pattern for this domain is centered on the invariant tryptophan.

Consensus pattern: [PSIAV]-x-[NDFV]-[NEQIY]-x-[LIVMAGP]-W-[NQSTHF]-[FYHQ]-[LIVMR]

[1] Boguski M.S., Murray A.W., Powers S. New Biol. 4:408-411(1992).

420. (PR55) Protein phosphatase 2A regulatory subunit PR55 signatures

Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric enzyme that consists of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit (PR65), also called subunit A; this complex then

associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme [1]. One of the forms of the variable subunit is a 55 Kd protein (PR55) which is highly conserved in mammals - where three isoforms are known to exist -, Drosophila and yeast (gene CDC55). This subunit could perform a substrate recognition function or be responsible for targeting the enzyme complex to the appropriate subcellular compartment.

As signature patterns, two perfectly conserved sequences of 15 residues were selected; one located in the N-terminal region, the other in the center of the protein.

Consensus pattern: E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N

Consensus pattern: N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D

[1] Mayer-Jaekel R., Hemmings B.A. Trends Cell Biol. 4:287-291(1994).

421. N-(5'phosphoribosyl)anthranilate (PRA) isomerase

[1] Wilmanns M, Priestle JP, Niermann T, Jansonius JN;
J Mol Biol 1992;223:477-507.

422. (PRK) Phosphoribulokinase signature

Phosphoribulokinase (EC 2.7.1.19) (PRK) [1,2] is one of the enzymes specific to the Calvin's reductive pentose phosphate cycle which is the major route by which carbon dioxide is assimilated and reduced by autotrophic organisms. PRK catalyzes the ATP-dependent phosphorylation of ribulose 5-phosphate into ribulose 1,5-bisphosphate which is the substrate for RubisCO.

PRK's of diverse origins show different properties with respect to the size of the protein, the subunit structure, or the enzymatic regulation. However an alignment of the sequences of PRK from plants, algae, photosynthetic and chemoautotrophic bacteria shows that there are a few regions of sequence similarity. As a signature pattern one of these regions was selected.

Consensus pattern: K-[LIVM]-x-R-D-x(3)-R-G-x-[ST]-x-E

[1] Kossmann J., Klintworth R., Bowien B. Gene 85:247-252(1989).

[2] Gibson J.L., Chen J.-H., Tower P.A., Tabita F.R. Biochemistry 29:8085-8093(1990).

423. (PRPP synt) Phosphoribosyl pyrophosphate synthetase signature

Phosphoribosyl pyrophosphate synthetase (EC 2.7.6.1) (PRPP synthetase) catalyzes the formation of PRPP from ATP and ribose 5-phosphate. PRPP is then used in various biosynthetic pathways, as for example in the formation of purines, pyrimidines, histidine and tryptophan. PRPP synthetase requires inorganic phosphate and magnesium ions for its stability and activity.

In mammals, three isozymes of PRPP synthetase are found; in yeast there are at least four isozymes.

As a signature pattern for this enzyme, a very conserved region was selected that has been suggested to be involved in binding divalent cations [1]. This region contains two conserved aspartic acid residues as well as a histidine, which are all potential ligands for a cation such as magnesium.

Consensus pattern: D-[LI]-H-[SA]-x-Q-[IMST]-[QM]-G-[FY]-F-x(2)-P-[LIVMFC]-D

[1] Bower S.G., Harlow K.W., Switzer R.L., Hoven-Jensen B. J. Biol. Chem. 264:10287-10291(1989).

424. (PRTP) Herpesvirus processing and transport protein

The members of this family are associate with capsid intermediates during packaging of the virus.

Number of members: 31

[1]

Medline: 98362148

Herpes simplex virus type 1 cleavage and packaging proteins UL15 and UL28 are associated with B but not C capsids during packaging. Yu D, Weller SK;

J Virol 1998;72:7428-7439.

425. Photosystem I psaG / psaK (PSI PSAK) proteins signature

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. It is found in the chloroplasts of plants and cyanobacteria. PSI is composed of at least 14 different subunits, two of which PSI-G (gene psaG) and PSI-K (gene psaK) are small hydrophobic proteins of about 7 to 9 Kd and evolutionary related [2]. Both seem to contain two transmembrane regions. Cyanobacteria seem to encode only for PSI-K.

As a signature pattern, the best-conserved region was selected which seems to correspond to the second transmembrane region.

-Consensus pattern: [GT]-F-x-[LIVM]-x-[DEA]-x(2)-[GA]-x-[GTA]-[SA]-x-G-H-x-[LIVM]-[GA]

[1] Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).

[2] Kjaerulff S., Andersen B., Nielsen V.S., Moller B.L., Okkels J.S. J. Biol. Chem. 268:18912-18916(1993).

426. PTR2 family proton/oligopeptide symporters signatures

A family of eukaryotic and prokaryotic proteins that seem to be mainly involved in the intake of small peptides with the concomitant uptake of a proton has been recently characterized [1,2]. Proteins that belong to this family are: - Fungal peptide transporter PTR2.

- Mammalian intestine proton-dependent oligopeptide transporter PeptT1.
- Mammalian kidney proton-dependent oligopeptide transporter PeptT2.
- Drosophila opt1.
- Arabidopsis thaliana peptide transporters PTR2-A and PTR2-B (also known as the histidine transporting protein NTR1).
- Arabidopsis thaliana proton-dependent nitrate/chlorate transporter CHL1.
- Lactococcus proton-dependent di- and tri-peptide transporter dtpT.
- Caenorhabditis elegans hypothetical protein C06G8.2.
- Caenorhabditis elegans hypothetical protein F56F4.5.

- Caenorhabditis elegans hypothetical protein K04E7.2.
- Escherichia coli hypothetical protein ybgH.
- Escherichia coli hypothetical protein ydgR.
- Escherichia coli hypothetical protein yhiP.
- Escherichia coli hypothetical protein yjdL.
- Bacillus subtilis hypothetical protein yclF.

These integral membrane proteins are predicted to comprise twelve transmembrane regions. As signature patterns, two of the best conserved regions were selected. The first is a region that includes the end of the second transmembrane region, a cytoplasmic loop as well as the third transmembrane region. The second pattern corresponds to the core of the fifth transmembrane region.

-Consensus pattern: [GA]-[GAS]-[LIVMFYWA]-[LIVM]-[GAS]-D-x-[LIVMFYWT]-[LIVMFYW]-G-x(3)-[TAV]-[IV]-x(3)-[GSTAV]-x-[LIVMF]-x(3)-[GA]
-Consensus pattern: [FYT]-x(2)-[LMFY]-[FYV]-[LIVMFYWA]-x-[IVG]-N-[LIVMAG]-G-[GSA]-[LIMF]

[1] Paulsen I.T., Skurray R.A. Trends Biochem. Sci. 19:404-404(1994).
[2] Steiner H.-Y., Naider F., Becker J.M. Mol. Microbiol. 16:825-834(1995).

427. Pumilio-family RNA binding domains (aka PUM-HD, Pumilio homology domain)

Puf domains are necessary and sufficient for sequence specific RNA binding in fly Pumilio and worm FBF-1 and FBF-2. Both proteins function as translational repressors in early embryonic development by binding sequences in the 3' UTR of target mRNAs (e.g. the nanos response element (NRE) in fly Hunchback mRNA, or the point mutation element (PME) in worm fem-3 mRNA). Other proteins that contain Puf domains are also plausible RNA binding proteins. JSN1_YEAST, for instance, appears to also contain a single RRM domain by HMM analysis.
Puf domains usually occur as a tandem repeat of 8 domains.
The Pfam model does not necessarily recognize all 8 domains in

all sequences; some sequences appear to have 5 or 6 domains on initial analysis, but further analysis suggests the presence of additional divergent domains.

[1] Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP. Nature 1997;390:477-484. [2] Zamore PD, Williamson JR, Lehmann R. RNA 1997;3:1421-1433.

428. PWWP domain. The PWWP domain is named after a conserved Pro-Trp-Trp-Pro motif. The function of the domain is currently unknown. Number of members: 19

[1] Medline: 98282232. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Stec I, Wright TJ, van Ommen GJB, de Boer PAJ, van Haeringen A, Moorman AFM, Altherr MR, den Dunnen JT; Hum Mol Genet 1998;7:1071-1082.

429. PX domain

Eukaryotic domain of unknown function present in phox proteins, PLD isoforms, a PI3K isoform.

Number of members: 71

[1]

Medline: 97084820

Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3-kinases: binding partners of SH3 domains?

Ponting CP;

Protein Sci 1996;5:2353-2357.

430. ParA family ATPase

[1]

Medline: 91141297

A family of ATPases involved in active partitioning of diverse bacterial plasmids.

Motallebi-Veshareh M, Rouch DA, Thomas CM;

Mol Microbiol 1990;4:1455-1463.

Number of members: 122

431. (Parvo coat) Parvovirus coat protein. 72 members.

432. Pectinesterase signatures

Pectinesterase (EC 3.1.1.11) (pectin methylesterase) catalyzes the hydrolysis of pectin into pectate and methanol. In plants, it plays an important role in cell wall metabolism during fruit ripening. In plant bacterial pathogens such as Erwinia carotovora and in fungal pathogens such as Aspergillus niger, pectinesterase is involved in maceration and soft-rotting of plant tissue. Prokaryotic and eukaryotic pectinesterases share a few regions of sequence similarity [1,2,3]. two of these regions were selected as signature patterns. The first is based on a region in the N-terminal section of these enzymes; it contains a conserved tyrosine which may play a role in the catalytic mechanism [3]. The second pattern corresponds to the best conserved region, an octapeptide located in the central part of these enzymes.

-Consensus pattern: [GSTNP]-x(6)-[FYVHR]-[IVN]-[KEP]-x-G-[STIVKRQ]-Y-[DNQKRMV]-[EP]-x(3)-[LIMVA]

-Consensus pattern: [IV]-x-G-[STAD]-[LIVT]-D-[FYI]-[IV]-[FSN]-G

[1] Ray J., Knapp J., Grierson D., Bird C., Schuch W. Eur. J. Biochem. 174:119-124(1988).

[2] Plastow G.S. Mol. Microbiol. 2:247-254(1988).

[3] Markovic O., Joernvall H. Protein Sci. 1:1288-1292(1992).

433. Pentapeptide repeats (8 copies)

These repeats are found in many cyanobacterial proteins.

The repeats were first identified in hglK [1]. The function of these repeats is unknown.

The structure of this repeat has been predicted to be a beta-helix [2].

The repeat can be approximately described as A(D/N)LXX, where X can be any amino acid.Number of members: 75

[1]

Medline: 96062225

The hglK gene is required for localization of heterocyst-specific glycolipids in the cyanobacterium Anabaena sp. strain PCC 7120.

Black K, Buikema WJ, Haselkorn R;

J Bacteriol 1995;177:6440-6448.

[2]Medline: 98318059

Structure and distribution of pentapeptide repeats in bacteria.

Bateman A, Murzin A, Teichmann SA;

Protein Sci 1998;7:1477-1480.

[3]Medline: 98316713

Characterisation of an Arabidopsis cDNA encoding a thylakoid lumen protein related to a novel 'pentapeptide repeat' family of proteins.

Kieselbach T, Mant A, Robinson C, Schroder WP;

FEBS Lett 1998;428:241-244.

434. Polypeptide deformylase

[1]

Medline: 97002011

A new subclass of the zinc metalloproteases superfamily revealed by the solution structure of peptide deformylase.

Meinnel T, Blanquet S, Dardel F;

J Mol Biol 1996;262:375-386.

[2]Medline: 98332750

Solution structure of nickel-peptide deformylase.
Dardel F, Ragusa S, Lazennec C, Blanquet S, Meinnel T;
J Mol Biol 1998;280:501-513.
Number of members: 21

435. Peptidyl-tRNA hydrolase signatures

Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH) is a bacterial enzyme that cleaves peptidyl-tRNA or N-acyl-aminoacyl-tRNA to yield free peptides or N-acyl-amino acids and tRNA. The natural substrate for this enzyme may be peptidyl-tRNA which drop off the ribosome during protein synthesis [1,2]. Bacterial PTH has been found [2,3] to be evolutionary related to yeast hypothetical protein YHR189w.

PTH and YHR189w are proteins of about 200 amino acid residues. As signature patterns, two conserved regions were selected that each contain an histidine. The first of these regions is located in the N-terminal section, the other in the central part.

-Consensus pattern: [FY]-x(2)-T-R-H-N-x-G-x(2)-[LIVMFA](2)-[DE]
-Consensus pattern: [GS]-x(3)-H-N-G-[LIVM]-[KR]-[DNS]-[LIVMT]

[1] Garcia-Villegas M.R., De La Vega F.M., Galindo J.M., Segura M., Buckingham R.H., Guarneros G. EMBO J. 10:3549-3555(1991).
[2] De La Vega F.M., Galindo J.M., Old I.G., Guarneros G. Gene 169:97-100(1996).
[3] Ouzounis C., Bork P., Casari G., Sander C. Protein Sci. 4:2424-2428(1995).

436. (Peptidase M17) Cytosol aminopeptidase signature

Cytosol aminopeptidase is a eukaryotic cytosolic zinc-dependent exopeptidase that catalyzes the removal of unsubstituted amino-acid residues from the N-terminus of proteins. This enzyme is often known as leucine aminopeptidase (EC 3.4.11.1) (LAP) but has been shown [1] to be identical with prolyl aminopeptidase (EC 3.4.11.5). Cytosol aminopeptidase is a hexamer of identical chains, each of which binds two zinc ions.

Cytosol aminopeptidase is highly similar to Escherichia coli pepA, a manganese dependent aminopeptidase. Residues involved in zinc ion-binding [2] in the mammalian enzyme are absolutely conserved in pepA where they presumably bind manganese.

A cytosol aminopeptidase from Rickettsia prowazekii [3] and one from Arabidopsis thaliana also belong to this family.

As a signature pattern for these enzymes, a perfectly conserved octapeptide was selected which contains two residues involved in binding metal ions: an aspartate and a glutamate.

-Consensus pattern: N-T-D-A-E-G-R-L [The D and the E are zinc/manganese ligands]

-Note: these proteins belong to family M17 in the classification of peptidases [4,E1].

[1] Matsushima M., Takahashi T., Ichinose M., Miki K., Kurokawa K., Takahashi K. Biochem. Biophys. Res. Commun. 178:1459-1464(1991).

[2] Burley S.K., David P.R., Sweet R.M., Taylor A., Lipscomb W.N. J. Mol. Biol. 224:113-140(1992).

[3] Wood D.O., Solomon M.J., Speed R.R. J. Bacteriol. 175:159-165(1993).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

437. Assemblin (Peptidase family S21)

[1]

Medline: 96399137

Three-dimensional structure of human cytomegalovirus protease.

Shieh HS, Kurumbail RG, Stevens AM, Stegeman RA, Sturman EJ, Pak JY, Wittwer AJ, Palmier MO, Wiegand RC, Holwerda BC, Stallings WC;

Nature 1996;383:279-282.

Number of members: 29

438. Pollen proteins Ole e I family signature

The following plant pollen proteins, whose biological function is not yet known, are structurally related [1]:

- Olive tree pollen major allergen (Ole e I).

- Tomato anther-specific protein LAT52. - Maize pollen-specific protein ZmC13.

These proteins are most probably secreted and consist of about 145 residues. As shown in the following schematic representation, there are six cysteines which are conserved in the sequence of these proteins. They seem to be involved in disulfide bonds.

xxxxxxCxCxxxxxxxxxCxxxxxxxxxxxxxxxxxCxxxxxCxxxxxxxxxxxxxxxxxxxCxxxxxxx

******'C': conserved cysteine involved in a disulfide bond.

'*': position of the pattern.

-Consensus pattern: [EQ]-G-x-V-Y-C-D-T-C-R [The two C's are probably involved in disulfide bonds]

[1] Villalba M., Batanero E., Lopez-Otin C., Sanchez L.M., Monsalve R.I., Gonzalez De La Pena M.A., Lahoz C., Rodriguez R. Eur. J. Biochem. 216:863-869(1993).

439. Pollen allergen

This family contains allergens lol PI, PII and PIII from Lolium perenne.

Number of members: 49

[1]

Medline: 90105394

Complete primary structure of a Lolium perenne (perennial rye grass) pollen allergen, Lol p III: comparison with known Lol p I and II sequences.

Ansari AA, Shenbagamurthi P, Marsh DG;

Biochemistry 1989;28:8665-8670.

440. Porphobilinogen deaminase cofactor-binding site

Porphobilinogen deaminase (EC 4.3.1.8), or hydroxymethylbilane synthase, is an enzyme involved in the biosynthesis of porphyrins and related macrocycles. It

catalyzes the assembly of four porphobilinogen (PBG) units in a head to tail fashion to form hydroxymethylbilane.

The enzyme covalently binds a dipyrromethane cofactor to which the PBG subunits are added in a stepwise fashion. In the Escherichia coli enzyme (gene hemC), this cofactor has been shown [1] to be bound by the sulfur atom of a cysteine. The region around this cysteine is conserved in porphobilinogen deaminases from various prokaryotic and eukaryotic sources.

-Consensus pattern: E-R-x-[LIVMFA]-x(3)-[LIVMF]-x-G-[GSA]-C-x-[IVT]-P-[LIVMF]-[GSA] [C is the cofactor attachment site]

[1] Miller A.D., Hart G.J., Packman L.C., Battersby A.R. Biochem. J. 254:915-918(1988).

441. Presenilin

Mutations in presenilin-1 are a major cause of early onset Alzheimer's disease [2]. It has been found that presenilin-1 (Swiss:P49768) binds to beta-catenin in vivo [4]. This family also contains SPE proteins from C.elegans.

Number of members: 23

[1]
Medline: 98045995
Presenilins and Alzheimer's disease.
Kim TW, Tanzi RE;
Curr Opin Neurobiol 1997;7:683-688.
[2]Medline: 98045995
Presenilins and Alzheimer's disease.
Kim TW, Tanzi RE;
Curr Opin Neurobiol 1997;7:683-688.
[3]Medline: 98099802
Interaction of presenilins with the filamin family of actin-binding proteins.
Zhang W, Han SW, McKeel DW, Goate A, Wu JY;
J Neurosci 1998;18:914-922.
[4]Medline: 99004850

Destabilisation of beta-catenin by mutations in presenilin-1 potentiates neuronal apoptosis.

Zhang Z, Hartmann H, Do VM, Abramowski D, Sturchler-Pierrat C, Staufenbiel M, Sommer B, van de Wetering M, Clevers H, Saftig P, De Strooper B, He X, Yankner BA;

Nature 1998;395:698-702.

442. (Pribosyltran) Purine/pyrimidine phosphoribosyl transferases signature

Phosphoribosyltransferases (PRT) are enzymes that catalyze the synthesis of beta-n-5'-monophosphates from phosphoribosylpyrophosphate (PRPP) and an enzyme specific amine. A number of PRT's are involved in the biosynthesis of purine, pyrimidine, and pyridine nucleotides, or in the salvage of purines and pyrimidines. These enzymes are:

- Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT), which is involved in purine salvage.
- Hypoxanthine-guanine or hypoxanthine phosphoribosyltransferase (EC 2.4.2.8) (HGPRT or HPRT), which are involved in purine salvage.
- Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT), which is involved in pyrimidine biosynthesis.
- Amido phosphoribosyltransferase (EC 2.4.2.14), which is involved in purine biosynthesis.
- Xanthine-guanine phosphoribosyltransferase (EC 2.4.2.22) (XGPRT), which is involved in purine salvage.

In the sequence of all these enzymes there is a small conserved region which may be involved in the enzymatic activity and/or be part of the PRPP binding site [1].

-Consensus pattern: [LIVMFYWCTA]-[LIVM]-[LIVMA]-[LIVMFC]-[DE]-D-[LIVMS]-[LIVM]-[STAVD]-[STAR]-[GAC]-x-[STAR]

-Note: in position 11 of the pattern most of these enzymes have Gly.

[1] Hershey H.V., Taylor M.W. Gene 43:287-293(1986).

443. (Pro CA)

Prokaryotic-type carbonic anhydrases signatures

Carbonic anhydrases (EC 4.2.1.1) (CA) are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide. In Escherichia coli, CA (gene cynT) is involved in recycling carbon dioxide formed in the bicarbonate-dependent decomposition of cyanate by cyanase (gene cynS). By this action, it prevents the depletion of cellular bicarbonate [1]. In photosynthetic bacteria and plant chloroplast, CA is essential to inorganic carbon fixation [2]. Prokaryotic and plant chloroplast CA are structurally and evolutionary related and form a family distinct from the one which groups the many different forms of eukaryotic CA's (see <PDOC00146>). Hypothetical proteins yadF from Escherichia coli and HI1301 from Haemophilus influenzae also belong to this family. Two signature patterns were developed for this family of enzymes. Both patterns contain conserved residues that could be involved in binding zinc (cysteine and histidine).

-Consensus pattern: C-[SA]-D-S-R-[LIVM]-x-[AP]

-Consensus pattern: [EQ]-Y-A-[LIVM]-x(2)-[LIVM]-x(4)-[LIVMF](3)-x-G-H-x(2)-C-G

[1] Guilloton M.B., Korte J.J., Lamblin A.F., Fuchs J.A., Anderson P.M. J. Biol. Chem. 267:3731-3734(1992).

[2] Fukuzawa H., Suzuki E., Komukai Y., Miyachi S. Proc. Natl. Acad. Sci. U.S.A. 89:4437-4441(1992).

444. (Prolyl_oligopep)

Prolyl oligopeptidase family serine active site

The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences.
- Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and argininyl residues.
- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.
- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.
- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).
- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase).

This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in E.coli proteaseII as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern: D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4].

[1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).

[2] Barrett A.J., Rawlings N.D.

[3] Polgar L., Szabo E.

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

445. (Pterin 4a)
Pterin 4 alpha carbinolamine dehydratase

Pterin 4 alpha carbinolamine dehydratase is aka DCoH (dimerisation cofactor of hepatocyte nuclear factor 1-alpha).

Number of members: 11

[1] Cronk JD, Endrizzi JA, Alber T; Medline: 97052967 High-resolution structures of the bifunctional enzyme and transcriptional coactivator DCoH and its complex with a product analogue." Protein Sci 1996;5:1963-1972.

446. (Pyridox oxidase)
Pyridoxamine 5'-phosphate oxidase signature

Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5) is a FMN flavoprotein involved in the de novo synthesis of pyridoxine (vitamin B6) and pyridoxal phosphate. It oxidizes pyridoxamine-5-P (PMP) and pyridoxine-5-P (PNP) to pyridoxal-5-P. The sequences of the enzyme from bacterial (genes pdxH or fprA) [1] and fungal (gene PDX3) [2] sources show that this protein has been highly conserved throughout evolution.
PdxH is evolutionary related [3] to one of the enzymes in the phenazine biosynthesis protein pathway, phzD (also known as phzG). As a signature pattern, a highly conserved region was selected located in the C-terminal part of these enzymes.

-Consensus pattern: [LIVF]-E-F-W-[QHG]-x(4)-R-[LIVM]-H-[DNE]-R

[1] Lam H.-M., Winkler M.E. J. Bacteriol. 174:6033-6045(1992).
[2] Loubbardi A., Karst F., Guilloton M., Marcireau C. J. Bacteriol. 177:1817-1823(1995).

[3] Pierson L.S. III, Gaffney T., Lam S., Gong F. FEMS Microbiol. Lett. 134:299-307(1995).

447. (Pyrophosphatase)
Inorganic pyrophosphatase signature

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) [1,2] is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed principally as the product of the many biosynthetic reactions that utilize ATP. All known Ppases require the presence of divalent metal cations, with magnesium conferring the highest activity. Among other residues, a lysine has been postulated to be part or close to the active site. PPases have been sequenced from bacteria such as Escherichia coli (homohexamer), thermophilic bacteria PS-3 and Thermus thermophilus, from the archaebacteria Thermoplasma acidophilum, from fungi (homodimer), from a plant, and from bovine retina. In yeast, a mitochondrial isoform of PPase has been characterized which seems to be involved in energy production and whose activity is stimulated by uncouplers of ATP synthesis.

The sequences of PPases share some regions of similarities. As signature patterns a region was selected that contains three conserved aspartates that are involved in the binding of cations.

-Consensus pattern: D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC]
[The three D's bind divalent metal cations]

[1] Lahti R., Kolakowski L.F. Jr., Heinonen J., Vihinen M., Pohjanoksa K., Cooperman B.S. Biochim. Biophys. Acta 1038:338-345(1990).
[2] Cooperman B.S., Baykov A.A., Lahti R. Trends Biochem. Sci. 17:262-266(1992).

448. (Peptidase S26)
Signal peptidases I signatures.

Signal peptidases (SPases) [1] (aka leader peptidases) remove the signal peptides from secretory proteins. In prokaryotes three types of SPasesare known: type I (gene lepB) which is responsible for the processing of the majority of exported pre-proteins; type II (gene lsp) which only process lipoproteins, and a third type involved in the processing of pili subunits. SPase I (EC 3.4.21.89) is an integral membrane protein that is anchored in the cytoplasmic membrane by one (in B. subtilis) or two (in E. coli) N-terminal transmembrane domains with the main part of the protein protuding in the periplasmic space. Two residues have been shown [2,3] to be essential for the catalytic activity of SPase I: a serine and an lysine. SPase I is evolutionary related to the yeast mitochondrial inner membrane protease subunit 1 and 2 (genes IMP1 and IMP2) which catalyze the removal of signal peptides required for the targeting of proteins from the mitochondrial matrix, across the inner membrane, into the inter-membrane space [4].In eukaryotes the removal of signal peptides is effected by an oligomeric enzymatic complex composed of at least five subunits: the signal peptidase complex (SPC). The SPC is located in the endoplasmic reticulum membrane. Two components of mammalian SPC, the 18 Kd (SPC18) and the 21 Kd (SPC21) subunits as well as the yeast SEC11 subunit have been shown [5] to share regions of sequence similarity with prokaryotic SPases I and yeast IMP1/IMP2. Three signature patterns have been developed for these proteins. The first signature contains the putative active site serine, the second signature contains the putative active site lysine which is not conserved in the SPC subunits, and the third signature corresponds to a conserved region of unknown biological significance which is located in the C-terminal section of all these proteins.

Consensus pattern: [GS]-x-S-M-x-[PS]-[AT]-[LF] [S is an active site residue]-
Consensus pattern: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY] [K is an active site residue]-
Consensus pattern: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG]-

[1] Dalbey R.E., von Heijne G. Trends Biochem. Sci. 17:474-478(1992).[2] Sung M., Dalbey R.E. J. Biol. Chem. 267:13154-13159(1992).[3] Black M.T. J. Bacteriol. 175:4957-4961(1993).[4] Nunnari J., Fox T.D., Walter P. Science 262:1997-2004(1993).[5] van Dijl J.M., de Jong A., Vehmaanpera J., Venema G., Bron S. EMBO J. 11:2819-2828(1992).[6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

449. (Peptidase C1) Eukaryotic thiol (cysteine) proteases active sites. Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences). - Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2]. - Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2]. - Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium- activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain. - Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3]. - Human cathepsin O [4]. - Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide). - Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A. - House-dust mites allergens DerP1 and EurM1. - Cathepsin B-like proteinases from the worms Caenorhabditis elegans (genes gcp-1, cpr-3, cpr-4, cpr-5 and cpr-6), Schistosoma mansoni (antigen SM31) and Japonica (antigen SJ31), Haemonchus contortus (genes AC-1 and AC-2), and Ostertagia ostertagi (CP-1 and CP-3). - Slime mold cysteine proteinases CP1 and CP2. - Cruzipain from Trypanosoma cruzi and brucei. - Throphozoite cysteine proteinase (TCP) from various Plasmodium species. - Proteases from Leishmania mexicana, Theileria annulata and Theileria parva. - Baculoviruses cathepsin-like enzyme (v-cath). - Drosophila small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain. - Yeast thiol protease BLH1/YCP1/LAP3. - Caenorhabditis elegans hypothetical protein C06G4.2, a calpain-like protein. Two bacterial peptidases are also part of this family: - Aminopeptidase C from Lactococcus lactis (gene pepC) [5]. - Thiol protease tpr from Porphyromonas gingivalis. Three other proteins are structurally related to this family, but may have lost their proteolytic activity. - Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine. - Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see <PDOC00382>). - Plasmodium falciparum serine-

repeat protein (SERA), the major blood stage antigen. This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine. The sequences around the three active site residues are well conserved and can be used as signature patterns.

Consensus pattern: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]- Note: the residue in position 4 of the pattern is almost always cysteine; the only exceptions are calpains (Leu), bleomycin hydrolase (Ser) and yeast YCP1 (Ser). -Note: the residue in position 5 of the pattern is always Gly except in papaya protease IV where it is Glu.

Consensus pattern: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH] [H is the active site residue]-

Consensus pattern: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue] - Note: these proteins belong to family C1 (papain-type) and C2 (calpains) in the classification of peptidases [7,E1].-

[1] Dufour E. Biochimie 70:1335-1342(1988).[2] Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).[3] Shi G.-P., Chapman H.A., Bhairi S.M., Deleeuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).[4] Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).[5] Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).[6] Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

450. (peptidase M24) Aminopeptidase P and proline dipeptidase signature (1). Aminopeptidase P (EC 3.4.11.9) is the enzyme responsible for the release of any N-terminal amino acid adjacent to a proline residue. Proline dipeptidase(EC 3.4.13.9) (prolidase) splits dipeptides with a prolyl residue in the carboxyl terminal position. Bacterial aminopeptidase P II (gene pepP) [1], proline dipeptidase (gene pepQ)[2], and human proline dipeptidase (gene PEPD) [3] are evolutionary related. These proteins are manganese metalloenzymes. Yeast hypothetical proteins YER078c and YFR006w and Mycobacterium tuberculosis hypothetical protein MtCY49.29c also belong to this family. As a signature pattern for these enzymes a conserved region that contains three histidine residues has been developed

Consensus pattern: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE]-

[1] Yoshimoto T., Tone H., Honda T., Osatomi K., Kobayashi R., Tsuru D. J. Biochem. 105:412-416(1989).[2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:6439-6439(1990).[3] Endo F., Tanoue A., Nakai H., Hata A., Indo Y., Titani K., Matsuda I. J. Biol. Chem. 264:4476-4481(1989).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

Methionine aminopeptidase signatures. (2). Methionine aminopeptidase (EC 3.4.11.18) (MAP) is responsible for the removal of the amino-terminal (initiator) methionine from nascent eukaryotic cytosolic and cytoplasmic prokaryotic proteins if the penultimate amino acid is small and uncharged. All MAP studied to date are monomeric proteins that require cobalt ions for activity. Two subfamilies of MAP enzymes are known to exist [1,2]. While being evolutionary related, they only share a limited amount of sequence similarity mostly clustered around the residues shown, in the Escherichia coli MAP [3],to be involved in cobalt-binding. The first family consists of enzymes from prokaryotes as well as eukaryoticMAP-1, while the second group is made up of archebacterial MAP and eukaryoticMAP-2. The second subfamily also includes proteins which do not seem to be MAP, but that are clearly evolutionary related such as mouse proliferation-associated protein 1 and fission yeast curved DNA-binding protein. For each of these subfamilies, a specific signature pattern that includes residues known to be involved in colbalt-binding has been developed.

Consensus pattern: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]- [YWV] [H is a cobalt ligand]-
Consensus pattern: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN] [The second D and the last D/N are cobalt ligands]

[1] Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E., Matthews B.W., Bradshaw R.A. Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).[2] Keeling P.J., Doolittle W.F. Trends Biochem. Sci. 21:285-286(1996).[3] Roderick S.L., Mathews B.W. Biochemistry 32:3907-3912(1993).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

451. Cytochrome P450 cysteine heme-iron ligand signature

Cytochrome P450's [1,2,3,E1] are a group of enzymes involved in the oxidative metabolism of a high number of natural compounds (such as steroids, fatty acids, prostaglandins, leukotrienes, etc) as well as drugs, carcinogens and mutagens. Based on sequence similarities, P450's have been classified into about forty different families [4,5]. P450's are proteins of 400 to 530 amino acids; the only exception is Bacillus BM-3 (CYP102) which is a protein of 1048residues that contains a N-terminal P450 domain followed by a reductase domain. P450's are heme proteins. A conserved cysteine residue in the C-terminal part of P450's is involved in binding the heme iron in the fifth coordination site. From a region around this residue, a ten residue signature was developed specific to P450's.

Consensus pattern: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD] [C is the heme iron ligand]-

[1] Nebert D.W., Gonzalez F.J. Annu. Rev. Biochem. 56:945-993(1987).

[2] Coon M.J., Ding X., Pernecky S.J., Vaz A.D.N. FASEB J. 6:669-673(1992).

[3] Guengerich F.P. J. Biol. Chem. 266:10019-10022(1991).

[4] Nelson D.R., Kamataki T., Waxman D.J., Guengerich F.P., Estrabrook R.W., Feyereisen R., Gonzalez F.J., Coon M.J., Gunsalus I.C., Gotoh O., Okuda K., Nebert D.W. DNA Cell Biol. 12:1-51(1993).

[5] Degtyarenko K.N., Archakov A.I. FEBS Lett. 332:1-8(1993).

452. (Pec Lyase) Pectate lyase

This enzyme forms a right handed beta helix structure. Pectate lyase is an enzyme involved in the maceration and soft rotting of plant tissue.

[1] Yoder MD, Keen NT, Jurnak F, Science 1993;260:1503-1507.

453. (pep M24) Aminopeptidase P and proline dipeptidase signature (pep1)

Aminopeptidase P (EC 3.4.11.9) is the enzyme responsible for the release of any N-terminal amino acid adjacent to a proline residue. Proline dipeptidase(EC 3.4.13.9) (prolidase) splits

dipeptides with a prolyl residue in the carboxyl terminal position. Bacterial aminopeptidase P II (gene pepP) [1], proline dipeptidase (gene pepQ)[2], and human proline dipeptidase (gene PEPD) [3] are evolutionary related. These proteins are manganese metalloenzymes. Yeast hypothetical proteins YER078c and YFR006w and Mycobacterium tuberculosis .hypothetical protein MtCY49.29c also belong to this family. As a signature pattern for these enzymes a conserved region was selected that contains three histidine residues.

Consensus pattern: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE]-

[1] Yoshimoto T., Tone H., Honda T., Osatomi K., Kobayashi R., Tsuru D. J. Biochem. 105:412-416(1989).

[2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:6439-6439(1990).

[3] Endo F., Tanoue A., Nakai H., Hata A., Indo Y., Titani K., Matsuda l. J. Biol. Chem. 264:4476-4481(1989).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

Methionine aminopeptidase signatures (pep2)

Methionine aminopeptidase (EC 3.4.11.18) (MAP) is responsible for the removal of the amino-terminal (initiator) methionine from nascent eukaryotic cytosolic and cytoplasmic prokaryotic proteins if the penultimate amino acid is small and uncharged. All MAP studied to date are monomeric proteins that require cobalt ions for activity. Two subfamilies of MAP enzymes are known to exist [1,2]. While being evolutionary related, they only share a limited amount of sequence similarity mostly clustered around the residues shown, in the Escherichia coli MAP [3],to be involved in cobalt-binding. The first family consists of enzymes from prokaryotes as well as eukaryotic MAP-1, while the second group is made up of archebacterial MAP and eukaryotic MAP-2. The second subfamily also includes proteins which do not seem to be MAP, but that are clearly evolutionary related such as mouse proliferation-associated protein 1 and fission yeast curved DNA-binding protein. For each of these subfamilies, a specific signature pattern was developed that includes residues known to be involved in colbalt-binding.

Consensus pattern: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]- [YWV] [H is a cobalt ligand]-

Consensus pattern: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN] [The second D and the last D/N are cobalt ligands]

[1] Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E., Matthews B.W., Bradshaw R.A. Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).

[2] Keeling P.J., Doolittle W.F. Trends Biochem. Sci. 21:285-286(1996).

[3] Roderick S.L., Mathews B.W. Biochemistry 32:3907-3912(1993).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

454. Peroxidases signatures

Peroxidases (EC 1.11.1.-) [1] are heme-binding enzymes that carry out a variety of biosynthetic and degradative functions using hydrogen peroxide as the electron acceptor. Peroxidases are widely distributed throughout bacteria, fungi, plants, and vertebrates. In peroxidases the heme prosthetic group is protoporphyrin IX and the fifth ligand of the heme iron is a histidine (known as the proximal histidine). Another histidine residue (the distal histidine) serves as an acid-base catalyst in the reaction between hydrogen peroxide and the enzyme. The regions around these two active site residues are more or less conserved in a majority of peroxidases [2,3]. The enzymes in which one or both of these regions can be found are listed below. - Yeast cytochrome c peroxidase (EC 1.11.1.5). - Myeloperoxidase (EC 1.11.1.7) (MPO). MPO is found in granulocytes and monocytes and plays a major role in the oxygen-dependent microbicidal system of neutrophils. - Lactoperoxidase (EC 1.11.1.7) (LPO). LPO is a milk protein which acts as an antimicrobial agent. - Eosinophil peroxidase (EC 1.11.1.7) (EPO). An enzyme found in the cytoplasmic granules of eosinophils. - Thyroid peroxidase (EC 1.11.1.8) (TPO). TPO plays a central role in the biosynthesis of thyroid hormones. It catalyzes the iodination and coupling of the hormonogenic tyrosines in thyroglobulin to yield the thyroid hormones T3 and T4. - Fungal ligninases. Ligninase catalyzes the first step in the degradation of lignin. It depolymerizes lignin by catalyzing the C(alpha)-C(beta) cleavage of the propyl side chains of lignin. - Plant peroxidases (EC 1.11.1.7). Plants expresses a large numbers of isozymes of peroxidases. Some of them play a role in cell-suberization by catalyzing the deposition of the aromatic residues of suberin on the cell wall, some are expressed as a defense response toward wounding, others are involved in the metabolism of auxin and the biosynthesis of lignin. - Prokaryotic catalase-peroxidases. Some bacterial species produce enzymes that exhibit both catalase and broad-spectrum

peroxidase activities [4]. Examples of such enzymes are: catalase HP I from Escherichia coli (gene katG) and perA from Bacillus stearothermophilus.

Consensus pattern: [DET]-[LIVMTA]-x(2)-[LIVM]-[LIVMSTAG]-[SAG]-[LIVMSTAG]-H-[STA]-[LIVMFY] [H is the proximal heme-binding ligand] -
Consensus pattern: [SGATV]-x(3)-[LIVMA]-R-[LIVMA]-x-[FW]-H-x-[SAC] [H is an active site residue]-

[1] Dawson J.H. Science 240:433-439(1988).
[2] Kimura S., Ikeda-Saito M. Proteins 3:113-120(1988).
[3] Henrissat B., Saloheimo M., Lavaitte S., Knowles J.K.C. Proteins 8:251-257(1990).
[4] Welinder K.G. Biochim. Biophys. Acta 1080:215-220(1991).

455. pfkB family of carbohydrate kinases signatures

It has been shown [1,2,3] that the following carbohydrate and purine kinasesare evolutionary related and can be grouped into a single family, which isknown [1] as the 'pfkB family': - Fructokinase (EC 2.7.1.4) (gene scrK). - 6-phosphofructokinase isozyme 2 (EC 2.7.1.11) (phosphofructokinase-2) (gene pfkB). pfkB is a minor phosphofructokinase isozyme in Escherichia coli and is not evolutionary related to the major isozyme (gene pfkA). Plants 6-phosphofructokinase also belong to this family. - Ribokinase (EC 2.7.1.15) (gene rbsK). - Adenosine kinase (EC 2.7.1.20) (gene ADK). - 2-dehydro-3-deoxygluconokinase (EC 2.7.1.45) (gene: kdgK). - 1-phosphofructokinase (EC 2.7.1.56) (fructose 1-phosphate kinase) (gene fruK). - Inosine-guanosine kinase (EC 2.7.1.73) (gene gsk). - Tagatose-6-phosphate kinase (EC 2.7.1.144) (phosphotagatokinase) (gene lacC). - Escherichia coli hypothetical protein yeiC. - Escherichia coli hypothetical protein yeiI. - Escherichia coli hypothetical protein yhfQ. - Escherichia coli hypothetical protein yihV. - Bacillus subtilis hypothetical protein yxdC. - Yeast hypothetical protein YJR105w.All the above kinases are proteins of from 280 to 430 amino acid residues that share a few region of sequence similarity. Two of these regions were selected as signature patterns. The first pattern is based on a region rich in glycine which is located in the N-terminal section of these enzymes; while the second pattern is based on a conserved region in the C-terminal section.

Consensus pattern: [AG]-G-x(0,1)-[GAP]-x-N-x-[STA]-x(6)-[GS]-x(9)-G-

Consensus pattern: [DNSK]-[PSTV]-x-[SAG](2)-[GD]-D-x(3)-[SAGV]-[AG]- [LIVMFYA]-[LIVMSTAP]

[1] Wu L.-F., Reizer A., Reizer J., Cai B., Tomich J.M., Saier M.H. Jr. J. Bacteriol. 173:3117-3127(1991).

[2] Orchard L.M.D., Kornberg H.L. Proc. R. Soc. Lond., B, Biol. Sci. 242:87-90(1990).

[3] Blatch G.L., Scholle R.R., Woods D.R. Gene 95:17-23(1990).

456. Phospholipase A2 active sites signatures

Phospholipase A2 (EC 3.1.1.4) (PA2) [1,2] is an enzyme which releases fatty acids from the second carbon group of glycerol. PA2's are small and rigid proteins of 120 amino-acid residues that have four to seven disulfide bonds.PA2 binds a calcium ion which is required for activity. The side chains of two conserved residues, a histidine and an aspartic acid, participate in a 'catalytic network'. Many PA2's have been sequenced from snakes, lizards, bees and mammals. In the latter, there are at least four forms: pancreatic, membrane-associated as well as two less characterized forms. The venom of most snakes contains multiple forms of PA2. Some of them are presynaptic neurotoxins which inhibit neuromuscular transmission by blocking acetylcholine release from the nerve termini. Two different signature patterns were derived for PA2's. The first is centered on the active site histidine and contains three cysteines involved in disulfide bonds. The second is centered on the active site aspartic acid and also contains three cysteines involved in disulfide bonds.

Consensus pattern: C-C-x(2)-H-x(2)-C [H is the active site residue] This pattern will not detect some snake toxins homologous with PA2 but which have lost their catalytic activity as well as otoconin-22, a Xenopus protein from the aragonitic otoconia which is also unlikely to be enzymatically active.

Consensus pattern: [LIVMA]-C-{LIVMFYWPCST}-C-D-x(5)-C [D is the active site residue] The majority of functional and non-functional PA2's. Undetected sequences are bee PA2, gila monster PA2's, PA2 PL-X from habu and PA2 PA-5 from mulga.

[1] Davidson F.F., Dennis E.A. J. Mol. Evol. 31:228-238(1990).

[2] Gomez F., Vandermeers A., Vandermeers-Piret M.-C., Herzog R., Rathe J., Stievenart M., Winand J., Christophe J. Eur. J. Biochem. 186:23-33(1989).

457. Phosphorylase pyridoxal-phosphate attachment site. Phosphorylases (EC 2.4.1.1) [1] are important allosteric enzymes in carbohydrate metabolism. They catalyze the formation of glucose 1-phosphatefrom polyglucose such as glycogen, starch or maltodextrin. Enzymes from different sources differ in their regulatory mechanisms and their natural substrates. However, all known phosphorylases share catalytic and structural properties. They are pyridoxal-phosphate dependent enzymes; the pyridoxal-P group is attached to a lysine residue around which the sequence is highly conserved and can be used as a signature pattern to detect this class of enzymes.

Consensus pattern: E-A-[SC]-G-x-[GS]-x-M-K-x(2)-[LM]-N [K is the pyridoxal-P attachment site]-

[1] Fukui T., Shimomura S., Nakano K. Mol. Cell. Biochem. 42:129-144(1982).

458. Protein kinases signatures and profile
Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common toboth serine/threonine and tyrosine protein kinases. There are a number ofconserved regions in the catalytic domain of protein kinases. Two of these regions were selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; Two signature patterns were derived for that region: one specific for serine/threonine kinases and the other for tyrosine kinases. A profile was also developed which is based on the alignment in [1] and covers the entire catalytic domain.

Consensus pattern: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]. The majority of known protein kinases belong to the class detected by this pattern, but

it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]. Most serine/ threonine specific protein kinases belong to this class detected by the pattern with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3) [D is an active site residue] ALL tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk belong to this class detected by the pattern. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses gangciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. This profile also detects receptor guanylate cyclases and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects Arabidopsis thaliana kinase- like protein TMKL1 which seems to have lost its catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus [11] and Yersinia pseudotuberculosis.

[1] Hanks S.K., Hunter T. FASEB J. 9:576-596(1995).

[2] Hunter T. Meth. Enzymol. 200:3-37(1991).

[3] Hanks S.K., Quinn A.M. Meth. Enzymol. 200:38-62(1991).

[4] Hanks S.K. Curr. Opin. Struct. Biol. 1:369-383(1991).

[5] Hanks S.K., Quinn A.M., Hunter T. Science 241:42-52(1988).

[6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor S.S., Sowadski J.M. Science 253:407-414(1991).

[7] Bairoch A., Claverie J.-M. Nature 331:22(1988).

[8] Benner S. Nature 329:21-21(1987).

[9] Kirby R. J. Mol. Evol. 30:489-492(1992).

[10] Littler E., Stuart A.D., Chee M.S. Nature 358:160-162(1992).

[11] Munoz-Dorado J., Inouye S., Inouye M. Cell 67:995-1006(1991).

Receptor tyrosine kinase class II signature

A number of growth factors stimulate mitogenesis by interacting with a familyof cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1]. These receptor tyrosine kinases (RTK)all share the same topology: an extracellular ligand-binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups. The prototype for class II RTK's is the insulin receptor, a heterotetramer of two alpha and two beta chains linked by disulfide bonds. The alpha and beta chains are cleavage products of a precursor molecule. The alpha chain contains the ligand binding site, the beta chain transverses the membrane and contains the tyrosine protein kinase domain. The receptors currently known to belong to class II are: - Insulin receptor from vertebrates. - Insulin growth factor I receptor from mammals. - Insulin receptor-related receptor (IRR), which is most probably a receptor for a peptide belonging to the insulin family. - Insects insulin-like receptors. - Molluscan insulin-related peptide(s) receptor (MIP-R). - Insulin-like peptide receptor from Branchiostoma lanceolatum. - The Drosophila developmental protein sevenless, a putative receptor for positional information required for the formation of the R7 photoreceptor cells. - The trk family of receptors (NTRK1, NTRK2 and NTRK3), which are high affinity receptors for nerve growth factor and related neurotrophic factors (BDNF and NT-3).And the following uncharacterized receptors: - ROS. - LTK (TYK1). - EDDR1 (cak, TRKE, RTK6). - NTRK3 (Tyro10, TKT). - A sponge putative receptor tyrosine kinase. While only the insulin and the insulin growth factor I receptors are known to exist in the tetrameric conformation specific to class II RTK's, all the above proteins share extensive homologies in their kinase domain, especially around the putative site of autophosphorylation. Hence, a signature pattern was developed for this class of RTK's, which includes the tyrosine residue, itself probably autophosphorylated.

Consensus pattern: [DN]-[LIV]-Y-x(3)-Y-Y-R [The second Y is the autophosphorylation site]

[1] Yarden Y., Ullrich A. Annu. Rev. Biochem. 57:443-478(1988).

Receptor tyrosine kinase class III signature

A number of growth factors stimulate mitogenesis by interacting with a family of cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1]. These receptor tyrosine kinases (RTK)all share the same topology: an extracellular ligand-

binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups. The class III RTK's are characterized by the presence of five to seven immunoglobulin-like domains [2] in their extracellular section. Their kinase domain differs from that of other RTK's by the insertion of a stretch of 70 to 100 hydrophilic residues in the middle ofthis domain. The receptors currently known to belong to class III are: - Platelet-derived growth factor receptor (PDGF-R). PDGF-R exists as a homo- or heterodimer of two related chains: alpha and beta [3]. - Macrophage colony stimulating factor receptor (CSF-1-R) (also known as the fms oncogene). - Stem cell factor (mast cell growth factor) receptor (also known as the kit oncogene). - Vascular endothelial growth factor (VEGF) receptors Flt-1 and Flk-1/KDR [4]. - Fl cytokine receptor Flk-2/Flt-3 [5]. - The putative receptor Flt-4 [7]. a signature pattern Was developed for this class of RTK's which is based on a conserved region in the kinase domain.

Consensus pattern: G-x-H-x-N-[LIVM]-V-N-L-L-G-A-C-T-

[1] Yarden Y., Ullrich A. Annu. Rev. Biochem. 57:443-478(1988).
[2] Hunkapiller T., Hood L. Adv. Immunol. 44:1-63(1989).
[3] Lee K.-H., Bowen-Pope D.F., Reed R.R. Mol. Cell. Biol. 10:2237-2246(1990).
[4] Terman B.I., Dougher-Vermazen M., Carrion M.E., Dimitrov D., Armellino D.C., Gospodarowicz D., Boehlen P. Biochem. Biophys. Res. Commun. 187:1579-1586(1992).
[5] Lyman S.D., James L., Vanden Bos T., de Vries P., Brasel K., Gliniak B., Hollingsworth L.T., Picha K.S., McKenna H.J., Splett R.R. Cell 75:1157-1167(1993).
[6] Galland F., Karamysheva A., Pebusque M.J., Borg J.P., Rottapel R., Dubreuil P., Rosnet O., Birnbaum D. Oncogene 8:1233-1240(1993).

Receptor tyrosine kinase class V signatures

A number of growth factors stimulate mitogenesis by interacting with a familyof cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1]. These receptor tyrosine kinases (RTK)all share the same topology: an extracellular ligand-binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups on the basis of sequence similarities. The extracellular domain of class V RTK's consist of a region of about 300amino acids, amongst which 16 conserved cysteines probably involved in disulfide bonds; this region is followed by two copies of a fibronectin typeIII domain. The ligands for these receptors are proteins of

about 200 to 300residues collectively known as Ephrins. The receptors currently known to belong to class V are [2,3,E1]: - EPHA1 (Eph-1; Esk). - EPHA2 (Eck; Mpk-5; Sek-2). - EPHA3 (Etk-1; Hek; Mek4; Tyro4; Rek4; Cek4). - EPHA4 (Sek; Hek8; Mpk-3; Cek8). - EPHA5 (Ehk-1; Hek7; Bsk; Cek7). - EPHA6 (Ehk-2). - EPHA7 (Ehk-3; Hek11; Mdk-1; Ebk). - EPHA8 (Eek). - EPHB1 (Eph-2; Elk; Net). - EPHB2 (Eph-3; Hek5; Drt; Erk; Nuk; Sek-3; Cek5; Qek5). - EPHB3 (Hek-2; Mdk-5). - EPHB4 (Htk; Mdk-2; Myk-1). - EPHB5 (Cek9).The EPHA subtype receptors bind to GPI-anchored ephrins while the EPHB subtype receptors bind to type-I membrane ephrins. Two signature patterns were developed for this class of RTK's, which each include some of the conserved cysteine residues.

Consensus pattern: F-x-[DN]-x-[GAW]-[GA]-C-[LIVM]-[SA]-[LIVM](2)-[SA]-[LV]-[KRHQ]-[LIVA]-x(3)-[KR]-C-[PSAW] [The two C's are probably involved in disulfide bonds]

Consensus pattern: C-x(2)-[DE]-G-[DEQ]-W-x(2,3)-[PAQ]-[LIVMT]-[GT]-x-C-x-C- x(2)-G-[HFY]-[EQ] [The three C's are probably involved in disulfide bonds]

[1] Yarden Y., Ullrich A. Annu. Rev. Biochem. 57:443-478(1988).

[2] Sajjadi F.G., Pasquale E.B., Subramani S. New Biol. 3:769-778(1991).

[3] Wicks I.P., Wilkinson D., Salvaris E., Boyd A.W. Proc. Natl. Acad. Sci. U.S.A. 89:1611-1615(1992).

459. Protein kinase C terminal domain

460. Plant thionins signature

Thionins are small, basic, plant proteins generally toxic to animal cells [1].They seem to exert their toxic effect at the level of the cell membrane but their exact function is not known. They consist of a polypeptide chain of fcrty five to fifty amino acids with three to four internal disulfide bonds. They are found in seeds but also in the cell wall of leaves [2]. Thionins are processed from larger precursor proteins [3]. Crambin [4], a hydrophobic plant seed protein, also belongs to this family. The pattern to detect this family of proteins includes three of the six cysteine residues involved in disulfide bonds. +----------------------------------+ |+---------------------------+ | || | | xxCCxxxxxxxxxxCxxxxxxxxxCxxxCxxCxxxxxCxxxxxxxx

************* | | | +---------+'C': conserved cysteine involved in a disulfide bond.'*': position of the pattern.

Consensus pattern: C-C-x(5)-R-x(2)-[FY]-x(2)-C [The three C's are involved in disulfide bonds] The proteins from the gamma-thionin family are not related to the above proteins and are described in a separate section.

[1] Vernon L.P., Evett G.E., Zeikus R.D., Gray W.R. Arch. Biochem. Biophys. 238:18-29(1985).
[2] Bohlmann H., Clausen S., Behnke S., Giese H., Hiller C., Reimann-Phillip U., Schrader G., Barkholt V., Apel K. EMBO J. 7:1559-1565(1988).
[3] Bohlmann H., Apel K. Mol. Gen. Genet. 207:446-454(1987).
[4] Teeter M.M., Mazer J.A., L'Italien J.J. Biochemistry 20:5437-5443(1981).

461. Polyprenyl synthetases signatures
A variety of isoprenoid compounds are synthesized by various organisms. For example in eukaryotes the isoprenoid biosynthetic pathway is responsible for the synthesis of a variety of end products including cholesterol, dolichol, ubiquinone or coenzyme Q. In bacteria this pathway leads to the synthesis of isopentenyl tRNA, isoprenoid quinones, and sugar carrier lipids. Among the enzymes that participate in that pathway, are a number of polyprenyl synthetase enzymes which catalyze a 1'4-condensation between 5 carbon isoprene units. Currently the sequence of some of these enzymes is known: - Eukaryotic farnesyl pyrophosphate synthetase (FPP synthetase) (EC 2.5.1.1 / EC 2.5.1.10) which catalyzes the sequential condensation of isopentenyl pyrophosphate (IPP) with dimethylallyl pyrophosphate (DMAPP), and then with the resultant geranyl pyrophosphate to form farnesyl pyrophosphate. FPP synthetase is a cytoplasmic dimeric enzyme. - Prokaryotic farnesyl pyrophosphate synthetase (gene ispA). - Prokaryotic octaprenyl diphosphate synthase (gene ispB). - Prokaryotic heptaprenyl diphosphate synthase (EC 2.5.1.30). - Eukaryotic geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (EC 2.5.1.1 / EC 2.5.1.10 / EC 2.5.1.29) which catalyzes the sequential addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate. In plants GGPP synthase is a chloroplast enzyme involved in the biosynthesis of terpenoids; in fungi, such as Neurospora crassa (gene al-3), this enzyme is involved in the biosynthesis of carotenoids. - Prokaryotic GGPP synthetase,

which are involved in the biosynthesis of carotenoids (gene crtE). Such an enzyme is also encoded in the cyanelle genome of Cyanophora paradoxa. - Eukaryotic hexaprenyl pyrophosphate synthetase, which is involved in the biosynthesis of coenzyme Q and which catalyzes the formation of all trans- polyprenyl pyrophosphates generally ranging in length of between 6 and 10 isoprene units depending on the species. HP synthetase is a mitochondrial membrane-associated enzyme. It has been shown [1 to 5] that all the above enzymes share some regions of sequence similarity. Two of these regions are rich in aspartic-acid residues and could be involved in the catalytic mechanism and/or the binding of the substrates. signature patterns were developed for both regions. Possible additional members of this family of proteins are: - Bacillus subtilis spore germination protein C3 (gene gerC3). Both proteins are most probably also enzymes involved in isoprenoid metabolism [6].

Consensus pattern: [LIVM](2)-x-D-D-x(2,4)-D-x(4)-R-R-[GH]-
Consensus pattern: [LIVMFY]-G-x(2)-[FYL]-Q-[LIVM]-x-D-D-[LIVMFY]-x-[DNG]

[1] Ashby M.N., Edwards P.A. J. Biol. Chem. 265:13157-13164(1990).
[2] Fujisaki S., Hara H., Nishimura Y., Horiuchi K., Nishino T. J. Biochem. 108:995-1000(1990).
[3] Carattoli A., Romano N., Ballario P., Morelli G., Macino G. J. Biol. Chem. 266:5854-5859(1991).
[4] Kuntz M., Roemer S., Suire C., Hugueney P., Weil J.H., Schantz R., Camara B. Plant J. 2:25-34(1992).
[5] Math S.K., Hearst J.E., Poulter C.D. Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764(1992).
[6] Bairoch A. Unpublished observations (1993).

462. Potato inhibitor I family signature

The potato inhibitor I family is one of the numerous families of serine proteinase inhibitors. Members of this protein family are found in plants; in the seeds of barley or beans [1,2,3], and in potato or tomato leaves where they accumulate in response to mechanical damage [4,5]. An inhibitor belonging to this family is also found in leech [6]. It is interesting to note that, currently, this is the only proteinase inhibitor family to be found both inplant and animal kingdoms. Structurally these inhibitors are small (60 to 90 residues) and in contrast with other families of protease inhibitors, they lack disulfide bonds. They have a single inhibitory

site. The consensus pattern includes three out of the four residues conserved in all members of this family and is located in the N-terminal half.

Consensus pattern: [FYW]-P-[EQH]-[LIV](2)-G-x(2)-[STAGV]-x(2)-A- Barley subtilisin-chymotrypsin inhibitor-2b has Glu instead of Gly. There is a trypsin inhibitor from the cucurbitaceae Momordica charantia [7], which is said to belong to the potato inhibitor I family but which shows only a very weak similarity with the other members of this family.

[1] Svendsen I., Hejgaard J., Chavan J.K. Carlsberg Res. Commun. 49:493-502(1984).
[2] Svendsen I., Boisen S., Hejgaard J. Carlsberg Res. Commun. 47:45-53(1982).
[3] Nozawa H., Yamagata H., Aizono Y., Yoshikawa M., Iwasaki T. J. Biochem. 106:1003-1008(1989).
[4] Cleveland T.E., Thornburg R.W., Ryan C.A. Plant Mol. Biol. 8:199-207(1987).
[5] Lee J.S., Brown W.E., Graham J.S., Pearce G., Fox E.A., Dreher T.W., Ahern K.G., Pearson G.D., Ryan C.A. Proc. Natl. Acad. Sci. U.S.A. 83:7277-7281(1986).
[6] Seemuller U., Eulitz M., Fritz H., Strobl A. Hoppe-Seyler's Z. Physiol. Chem. 361:1841-1846(1980).
[7] Zeng F.-Y., Qian R.-Q., Wang Y. FEBS Lett. 234:35-38(1988).

463. (pp binding) Phosphopantetheine attachment site

Phosphopantetheine (or pantetheine 4' phosphate) is the prosthetic group of acyl carrier proteins (ACP) in some multienzyme complexes where it serves as a 'swinging arm' for the attachment of activated fatty acid and amino-acid groups [1]. Phosphopantetheine is attached to a serine residue in these proteins [2]. ACP proteins or domains have been found in various enzyme systems which are listed below (references are only provided for recently determined sequences). - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to the different enzymatic activities; ACP is one of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the ACP domain is located in the N-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional enzyme; the ACP domain is located between the beta-ketoacyl reductase domain and the C-terminal thioesterase domain [3]. - Polyketide antibiotics synthase enzyme systems. Polyketides are secondary metabolites produced from

simple fatty acids, by microorganisms and plants. ACP is one of the polypeptidic components involved in the biosynthesis of Streptomyces polyketide antibiotics actinorhodin, curamycin, granatacin, monensin, oxytetracycline and tetracenomycin C. - Bacillus subtilis putative polyketide synthases pksK, pksL and pksM which respectively contain three, five and one ACP domains. - The multifunctional 6-methysalicylic acid synthase (MSAS) from Penicillium patulum. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which contains an ACP domain in the C-terminal extremity. - Multifunctional mycocerosic acid synthase (gene mas) from Mycobacterium bovis. - Gramicidin S synthetase I (gene grsA) from Bacillus brevis. This enzyme catalyzes the first step in the biosynthesis of the cyclic antibiotic gramicidin S. - Tyrocidine synthetase I (gene tycA) from Bacillus brevis. The reaction carried out by tycA is identical to that catalyzed by grsA - Gramicidin S synthetase II (gene grsB) from Bacillus brevis. This enzyme is a multifunctional protein that activates and polymerizes proline, valine, ornithine and leucine. GrsB contains four ACP domains. - Erythronolide synthase proteins 1, 2 and 3 from Saccharopolyspora erythraea which is involved in the biosynthesis of the polyketide antibiotic erythromicin. Each of these proteins contain two ACP domains. - Conidial green pigment synthase from Aspergillus nidulans. - ACV synthetase from various fungi. This enzyme catalyzes the first step in the biosynthesis of penicillin and cephalosporin. It contains three ACP domains. - Enterobactin synthetase component F (gene entF) from Escherichia coli. This enzyme is involved in the ATP-dependent activation of serine during enterobactin (enterochelin) biosynthesis. - Cyclic peptide antibiotic surfactin synthase subunits 1, 2 and 3 from Bacillus subtilis. Subunits 1 and 2 contains three related domains while subunit 3 only contains a single domain. - HC-toxin synthetase (gene HTS1) from Cochliobolus carbonum. This enzyme synthesizes HC-toxin, a cyclic tetrapeptide. HTS1 contains four ACP domains. - Fungal mitochondrial ACP [9], which is part of the respiratory chain NADH dehydrogenase (complex I). - Rhizobium nodulation protein nodF, which probably acts as an ACP in the synthesis of the nodulation Nod factor fatty acyl chain.The sequence around the phosphopantetheine attachment site is conserved in all these proteins and can be used as a signature pattern. A profile was also developed that spans the complete ACP-like domain.

Consensus pattern: [DEQGSTALMKRH]-[LIVMFYSTAC]-[GNQ]-[LIVMFYAG]-[DNEKHS]-S- [LIVMST]-{PCFY}-[STAGCPQLIVMF]-[LIVMATN]-[DENQGTAKRHLM]- [LIVMWSTA]-[LIVGSTACR]-x(2)-[LIVMFA] [S is the pantetheine attachment site]

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

[2] Pugh E.L., Wakil S.J. J. Biol. Chem. 240:4727-4733(1965).

[3] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991).

[6] Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G., Galizzi A., Albertini A.M. Gene 130:65-71(1993).

[9] Sackmann U., Zensen R., Rohlen D., Jahnke U., Weiss H. Eur. J. Biochem. 200:463-469(1991).

464. (Prenyltrans) Terpene synthases signature

The following enzymes catalyze mechanistically related reactions which involvethe highly complex cyclic rearrangement of squalene or its 2,3 oxide: - Lanosterol synthase (EC 5.4.99.7) (oxidosqualene--lanosterol cyclase), which catalyzes the cyclization of (S)-2,3-epoxysqualene to lanosterol, the initial precursor of cholesterol, steroid hormones and vitamin D in vertebrates and of ergosterol in fungi (gene ERG7). - Cycloartenol synthase (EC 5.4.99.8) (2,3-epoxysqualene--cycloartenol cyclase), a plant enzyme that catalyzes the cyclization of (S)-2,3- epoxysqualene to cycloartenol. - Hopene synthase (EC 5.4.99.-) (squalene--hopene cyclase), a bacterial enzyme that catalyzes the cyclization of squalene into hopene, a key step in hopanoid (triterpenoid) metabolism.These enzymes are evolutionary related [1] proteins of about 70 to 85 Kd. As a signature pattern, a highly conserved region was selected which is rich in aromatic residues and which is located in the C-terminal section.

Consensus pattern: [DE]-G-S-W-x-G-x-W-[GA]-[LIVM]-x-[FY]-x-Y-[GA]

[1] Corey E.J., Matsuda S.P.T., Bartel B. Proc. Natl. Acad. Sci. U.S.A. 90:11628-11632(1993).

465. Prion protein signatures

Prion protein (PrP) [1,2,3] is a small glycoprotein found in high quantity in the brains of humans or animals infected with a number of degenerative neurological diseases such as

Kuru, Creutzfeldt-Jacob disease (CJD), scrapie or bovine spongiform encephalopathy (BSE). PrP is encoded in the host genome and expressed both in normal and infected cells. It has a tendency to aggregate yielding polymers called rods. Structurally, PrP is a protein consisting of a signal peptide, followed by an N-terminal domain that contains tandem repeats of a short motif (PHGGGWGQin mammals, PHNPGY in chicken), itself followed by a highly conserved domain lly comes a C-terminal hydrophobic domain post-translationally removed when PrP is attachedto the extracellular side of the cell membrane by a GPI-anchor. The structureof PrP is shown in the following schematic representation: +---+----------------+- ******-------------------****-----+-----+ |Sig| Tandem repeats | C C S| | +---+----------------+--- -----------------|--------|----|+-----+ +--------+ | GPI'C': conserved cysteine involved in a disulfide bond.'*': position of the patterns. As signature pattern for PrP, a perfectly conserved alanine- and glycine-rich region of 16 residues was selected as well as a region centered on the second cysteine involved in the disulfide bond.

Consensus pattern: A-G-A-A-A-A-G-A-V-V-G-G-L-G-G-Y-
Consensus pattern: E-x-[ED]-x-K-[LIVM](2)-x-[KR]-[LIVM](2)-x-[QE]-M-C-x(2)- Q-Y [C is involved in a disulfide bond]

[1] Stahl N., Prusiner S.B. FASEB J. 5:2799-2807(1991).
[2] Brunori M., Chiara Silvestrini M., Pocchiari M. Trends Biochem. Sci. 13:309-313(1988).
[3] Prusiner S.B. Annu. Rev. Microbiol. 43:345-374(1989).

466. Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature and profile (pro isomerase)

Cyclophilin [1] is the major high-affinity binding protein in vertebrates for the immunosuppressive drug cyclosporin A (CSA). It exhibits a peptidyl- prolyl cis-trans isomerase activity (EC 5.2.1.8) (PPIase or rotamase). PPIase is an enzyme that accelerates protein folding by catalyzing the cis-transisomerization of proline imidic peptide bonds in oligopeptides [2]. It is probable that CSA mediates some of its effects via an inhibitory action on PPIase. Cyclophilin is a cytosolic protein which belongs to a family [3,4,5]that also includes the following isozymes: - Cyclophilin B (or S-cyclophilin), a PPIase which is retained in an endoplasmic reticulum compartment. - Cyclophilin C, a cytoplasmic PPiase. - Mitochondrial matrix cyclophilin (cyp3). - A PPIase which seems specific for the folding of

rhodopsin and is an integral membrane protein anchored by a C-terminal transmembrane region. This protein was first characterized in Drosophila (gene ninaA). - Bacterial periplasmic PPiase (gene ppiA). - Bacterial cytosolic PPiase (gene ppiB). - Natural-killer cell cyclophilin-related protein. This large protein (about 160 Kd) is a component of a putative tumor-recognition complex involved in the function of NK cells. It contains a cyclophilin-type PPiase domain. - Mammalian nucleoporin Nup358 [6], a nuclear pore complex protein of 358 Kd that contains a C-terminal cyclophilin-type PPiase domain. - Yeast hypothetical protein YJR032w. - Fission yeast hypothetical protein SpAC21E11.05c. - Caenorhabditis elegans hypothetical protein T27D1.1.The sequences of the different forms of cyclophilin-type PPIases are well conserved. As a signature pattern, a conserved region was selected in the central part of these enzymes.

Consensus pattern: [FY]-x(2)-[STCNLV]-x-F-H-[RH]-[LIVMN]-[LIVM]-x(2)-F- [LIVM]-x-Q-[AG]-G- FKBP's, a family of proteins that bind the immunosuppressive drug FK506, are also PPIases, but their sequence is not at all related to that of cyclophilin.

[1] Stamnes M.A., Rutherford S.L., Zuker C.S. Trends Cell Biol. 2:272-276(1992).
[2] Fischer G., Schmid F.X. Biochemistry 29:2205-2212(1990).
[3] Trandinh C.C., Pao G.M., Saier M.H. Jr. FASEB J. 6:3410-3420(1992).
[4] Galat A. Eur. J. Biochem. 216:689-707(1993).
[5] Hacker J., Fischer G. Mol. Microbiol. 10:445456(1993).
[6] Wu J., Matunis M.J., Kraemer D., Blobel G., Coutavas E. J. Biol. Chem. 270:14209-14213(1995).

467. Profilin signature

Profilin [1,2] is a small eukaryotic protein that binds to monomeric actin(G-actin) in a 1:1 ratio thus preventing the polymerization of actin into filaments (F-actin). It can also, in certain circumstance promotes actin polymerization. Profilin also binds to polyphosphoinositides such as PIP2.Overall sequence similarity among profilin from organisms which belong to different phyla (ranging from fungi to mammals) is low, but the N-terminal region is relatively well conserved. That region is thought to be involved inthe binding to actin. The signature pattern for profilin is based on conserved residues at the N-

terminal extremity .A protein structurally similar to profilin is present in the genome of variola and vaccinia viruses (gene A42R).

Consensus pattern: <x(0,1)-[STA]-x(0,1)-W-[DENQH]-x-[YI]-x-[DEQ]

[1] Haarer B.K., Brown S.S. Cell Motil. Cytoskeleton 17:71-74(1990).
[2] Sohn R.H., Goldschmidt-Clermont P. BioEssays 16:465-472(1994).

468. Protamine P1 signature

Protamines are small, highly basic proteins, that substitute for histones in sperm chromatin during the haploid phase of spermatogenesis. They pack sperm DNA into a highly condensed, stable and inactive complex. There are two different types of mammalian protamine, called P1 and P2. P1 has been found in all species studied, while P2 is sometimes absent. There seems to be a single type of avian protamine whose sequence is closely related to that of mammalian P1 [1].As a signature for this family of proteins, a conserved region was selected at the N-terminal extremity of the sequence.

Consensus pattern: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S-

[1] Oliva R., Goren R., Dixon G.H. J. Biol. Chem. 264:17627-17630(1989).

469. Sperm histone P2 (protamine P2)

This protein also known as protamine P2 can substitute for histones in the chromatin of sperm. The alignment contains both the sequence of the mature P2 protein and its propeptide.

470. Proteasome A-type subunits signature

The proteasome (or macropain) (EC 3.4.99.46) [1 to 5,E1] is an eukaryotic and archaebacterial multicatalytic proteinase complex that seems to be involved inan ATP/ubiquitin-dependent nonlysosomal proteolytic pathway. In eukaryotes the proteasome is composed of about 28 distinct subunits which form a highly ordered ring-shaped structure (20S ring) of about 700 Kd. Most proteasome subunits can be classified, on the basis on

sequence similarities into two groups, A and B. Subunits that belong to the A-type group are proteins of from 210 to 290 amino acids that share a number of conserved sequence regions. Subunits that are known to belong to this family are listed below. - Vertebrate subunits C2 (nu), C3, C8, C9, iota and zeta. - Drosophila PROS-25, PROS-28.1, PROS-29 and PROS-35. - Yeast C1 (PRS1), C5 (PRS3), C7-alpha (Y8) (PRS2), Y7, Y13, PRE5, PRE6 and PUP2. - Arabidopsis thaliana subunits alpha and PSM30. - Thermoplasma acidophilum alpha-subunit. In this archaebacteria the proteasome is composed of only two different subunits.As a signature pattern for proteasome A-type subunits the best conserved region was selected, which is located in the N-terminal part of these proteins.

Consensus pattern: [FY]-x(4)-[STNV]-x-[FYW]-S-P-x-G-[RKH]-x(2)-Q-[LIVM]-[DE]- Y-[SAD]-x(2)-[SAG]-. These proteins belong to family T1 in the classification of peptidases [6,E2].

[1] Rivett A.J. Biochem. J. 291:1-10(1993).
[2] Rivett A.J. Arch. Biochem. Biophys. 268:1-8(1989).
[3] Goldberg A.L., Rock K.L Nature 357:375-379(1992).
[4] Wilk S. Enzyme Protein 47:187-188(1993).
[5] Hilt W., Wolf D.H. Trends Biochem. Sci. 21:96-102(1996).
[6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

Proteasome B-type subunits signature

The proteasome (or macropain) (EC 3.4.99.46) [1 to 5,E1] is an eukaryotic and archaebacterial multicatalytic proteinase complex that seems to be involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway. In eukaryotes the proteasome is composed of about 28 distinct subunits which form a highly ordered ring-shaped structure (20S ring) of about 700 Kd. Most proteasome subunits can be classified, on the basis on sequence similarities into two groups, A and B. Subunits that belong to the B-type group are proteins of from 190 to 290 amino acids that share a number of conserved sequence regions. Subunits that are known to belong to this family are listed below. - Vertebrate subunits C5, beta, delta, epsilon, theta (C10-II), LMP2/RING12, C13 (LMP7/RING10), C7-I and MECL-1. - Yeast PRE1, PRE2 (PRG1), PRE3, PRE4, PRS3, PUP1 and PUP3. - Drosophila L(3)73AI. - Fission yeast pts1. - Thermoplasma acidophilum beta-subunit. In this archaebacteria the proteasome is composed of only two different subunits. As a signature

pattern for proteasome B-type subunits the best conserved region was selected, which is located in the N-terminal part of these proteins.

Consensus pattern: [LIVMA]-[GSA]-[LIVMF]-x-[FYLVGAC]-x(2)-[GSACFY]-[LIVMSTAC](3)-[GAC]-[GSTACV]-[DES]-x(15)-[RK]-x(12,13)-G-x(2)-[GSTA]-D-. These proteins belong to family T1 in the classification of peptidases [6,E2].

[1] Rivett A.J. Biochem. J. 291:1-10(1993).
[2] Rivett A.J. Arch. Biochem. Biophys. 268:1-8(1989).
[3] Goldberg A.L., Rock K.L Nature 357:375-379(1992).
[4] Wilk S. Enzyme Protein 47:187-188(1993).
[5] Hilt W., Wolf D.H. Trends Biochem. Sci. 21:96-102(1996).
[6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

471. (pyr redox) Pyridine nucleotide-disulphide oxidoreductases class-I active site
The pyridine nucleotide-disulphide oxidoreductases are FAD flavoproteins which contains a pair of redox-active cysteines involved in the transfer of reducing equivalents from the FAD cofactor to the substrate. On the basis of sequence and structural similarities [1] these enzymes can be classified into two categories. The first category groups together the following enzymes [2 to 6]: - Glutathione reductase (EC 1.6.4.2) (GR). - Higher eukaryotes thioredoxin reductase (EC 1.6.4.5). - Trypanothione reductase (EC 1.6.4.8). - Lipoamide dehydrogenase (EC 1.8.1.4), the E3 component of alpha-ketoacid dehydrogenase complexes. - Mercuric reductase (EC 1.16.1.1).The sequence around the two cysteines involved in the redox-active disulfide bond is conserved and can be used as a signature pattern.

Consensus pattern: G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P [The two C's form the active site disulfide bond]. In positions 6 and 7 of the pattern all known sequences have Asn-(Val/ Ile) with the exception of GR from plant chloroplasts and from cyanobacteria which have Ile-Arg [7].

[1] Kurlyan J., Krishna T.S.R., Wong L., Guenther B., Pahler A., Williams C.H. Jr., Model P. Nature 352:172-174(1991).
[2] Rice D.W., Schulz G.E., Guest J.R. J. Mol. Biol. 174:483-496(1984).

[3] Brown N.L. Trends Biochem. Sci. 10:400-402(1985).

[4] Carothers D.J., Pons G., Patel M.S. Arch. Biochem. Biophys. 268:409-425(1989).

[5] Walsh C.T., Bradley M., Nadeau K. Trends Biochem. Sci. 16:305-309(1991).

[6] Gasdaska P.Y., Gasdaska J.R., Cochran S., Powis G. FEBS Lett. 373:5-9(1995).

[7] Creissen G., Edwards E.A., Enard C., Wellburn A., Mullineaux P. Plant J. 2:129-131(1991).

472. (pyridoxal deC) DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site (pyridoxal deC)

Three different enzymes - all pyridoxal-dependent decarboxylases - seem to share regions of sequence similarity [1,2,3,4], especially in the vicinity of the lysine residue which serves as the attachment site for the pyridoxal-phosphate (PLP) group. These enzymes are: - Glutamate decarboxylase (EC 4.1.1.15) (GAD). Catalyzes the decarboxylation of glutamate into the neurotransmitter GABA (4-aminobutanoate). - Histidine decarboxylase (EC 4.1.1.22) (HDC). Catalyzes the decarboxylation of histidine to histamine. There are two completely unrelated types of HDC: those that use PLP as a cofactor (found in Gram-negative bacteria and mammals), and those that contain a covalently bound pyruvoyl residue (found in Gram-positive bacteria). - Aromatic-L-amino-acid decarboxylase (EC 4.1.1.28) (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase. DDC catalyzes the decarboxylation of tryptophan to tryptamine. It also acts on 5-hydroxy- tryptophan and dihydroxyphenylalanine (L-dopa). - Tyrosine decarboxylase (EC 4.1.1.25) (TyrDC) which converts tyrosine into tyramine, a precursor of isoquinoline alkaloids and various amides.These enzymes are collectively known as group II decarboxylases [3,4].

Consensus pattern: S-[LIVMFYW]-x(5)-K-[LIVMFYWG](2)-x(3)-[LIVMFYW]-x-[CA]-x(2)-[LIVMFYWQ]-x(2)-[RK] [K is the pyridoxal-P attachment site]

[1] Jackson F.R. J. Mol. Evol. 31:325-329(1990).

[2] Joseph D.R., Sullivan P., Wang Y.-M., Kozak C., Fenstermacher D.A., Behrendsen M.E., Zahnow C.A. Proc. Natl. Acad. Sci. U.S.A. 87:733-737(1990).

[3] Sandmeier E., Hale T.I., Christen P. Eur. J. Biochem. 221:997-1002(1994).

[4] Ishii S., Mizugichi H., Nishino J., Hayashi H., Kagamiyama H. J. Biochem. 120:369-376(1996).

473. Regulator of chromosome condensation (RCC1) signatures (RCC1)

The regulator of chromosome condensation (RCC1) [1] is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein, to promote the loss of bound GDP and the uptake offresh GTP, thus acting as a guanine-nucleotide dissociation stimulator (GDS)[2]. The interaction of RCC1 with ran probably plays an important role in the regulation of gene expression. RCC1, known as PRP20 or SRM1 in yeast, pim1 in fission yeast and BJ1 in Drosophila, is a protein that contains seven tandem repeats of a domain of about 50 to 60 amino acids. As shown in the following schematic representation, the repeats make up the major part of the length of the protein. Outside the repeat region, there is just a small N-terminal domain of about 40 to 50 residues and, in the Drosophila protein only, a C-terminal domain of about 130 residues. +----+-------+-------+-------+-------+-------+-------+-------+-------------+ |N-t.|Rpt. 1 |Rpt. 2 |Rpt. 3 |Rpt. 4 |Rpt. 5 |Rpt. 6 |Rpt. 7 | C-terminal | +----+-------+-------+-------+-------+-------+-------+-------+-------------+ In Drosophila two signature patterns for RCC1 were developed. The first is found in the N- terminal part of the second repeat; this is the most conserved part of RCC1. The second is derived from conserved positions in the C-terminal part of each repeat and detects up to five copies of the repeated domain. The RCC1-type of repeat is also found in the X-linked retinitis pigmentosa GTPase regulator [3].

Consensus pattern: G-x-N-D-x(2)-[AV]-L-G-R-x-T-

Consensus pattern: [LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM]-

[1] Dasso M. Trends Biochem. Sci. 18:96-101(1993).

[2] Boguski M.S., McCormick F. Nature 366:643-654(1993).

[3] Roepman R., Van Duijnhoven G., Rosenberg T., Pinckers A.J.L.G., Bleeker-Wagemakers L.M., Bergen A.A.B., Post J., Beck A., Reinhardt R., Ropers H.-H., Cremers F., Berger W. Hum. Mol. Genet. 5:1035-1041(1996).

474. RNA 3'-terminal phosphate cyclase signature (RCT)

RNA 3'-terminal phosphate cyclase (EC <u>6.5.1.4</u>) [1,2] catalyzes the conversion of 3'-phosphate to a 2',3'-cyclic phosphodiester at the end of RNA. The biological role of this

enzyme is unknown but it is likely to function in some aspects of cellular RNA processing. The reaction catalyzed by the enzyme occurs in three steps: 1) adenylation of the enzyme by ATP; 2) the enzyme acts on RNA-3'terminal phosphate to produce RNA-3'terminal diphosphate adenylate; 3) Release of AMP and cyclisation by a non catalytic nucleophilic attack by the adjacent 2'hydroxyl on the phosphorus in the diester linkage. This enzyme, which has been characterized in human (where there seems to be at least three isozymes) and Escherichia coli (gene rtCA), seems to be taxonomically widespread. It is found in insects, plants, fungi (gene RTC1 inyeast) and in archeabacteria. RNA cyclase is a protein of from 36 to 42 Kd. The best conserved region, which is used as a signature pattern, is a glycine-rich stretch of residues located in the central part of the sequence and which is reminiscent of various ATP, GTPor AMP glycine-rich loops. In this context, the conserved Arg (His in the E.coli enzyme) could be the AMP-binding residue.

Consensus pattern: [RH]-G-x(2)-P-x-G(3)-x-[LIV]-

[1] Genschik P., Billy E., Swianiewicz M., Filipowicz W. EMBO J. 16:2955-2967(1997).
[2] Filipowicz W., Vincente O. Meth. Enzymol. 181:499-510(1990).

475. REV protein (anti-repression trans-activator protein)

476. Prokaryotic-type class I peptide chain release factors signature (RF-1)

Peptide chain release factors (RFs) are required for the termination of protein biosynthesis [1]. At present two classes of RFs can be distinguished. Class I RFs bind to ribosomes that have encountered a stop codon at their decoding site and induce release of the nascent polypeptide. Class II RFs are GTP-binding proteins that interact with class I RFs and enhance class I RF activity. In prokaryotes there are two class I RFs that act in a codon specific manner[2]: RF-1 (gene prfA) mediates UAA and UAG-dependent termination while RF-2(gene prfB) mediates UAA and UGA-dependent termination. RF-1 and RF-2 are structurally and evolutionary related proteins which have been shown [3] to make up a family that also contains the following proteins: - Fungal MRF1, a mitochondrial RF (m-RF) which recognizes the UAA and UAG codons. - Escherichia coli RF-H, a protein of unknown function. - Escherichia coli hypothetical protein yaeJ and a close Pseudomonas putida

homolog. A highly conserved region located in the central part of the 40 to 45 Kd RF-1/2 and m-RF and in the N-terminal of the 15 to 16Kd RF-H and yaeJ is used as a signature pattern.

Consensus pattern: [AR]-[STA]-x-G-x-G-G-Q-[HNGCS]-V-N-x(3)-[ST]-A-[IV]
Note that prokaryotic-type class I RFs display no significant sequence similarity to prokaryotic-type class II which belong to the family of GTP-binding elongation factors nor to eukaryotic class I or class II RFs.

[1] Tate W.P. , Poole E.S., Mannering S.M. Prog. Nucleic Acids. Res. Mol. Biol. 52:293-335(1996).
[2] Craigen W.J., Lee C.C., Caskey C.T. Mol. Microbiol. 4:861-865(1990).
[3] Pel H.J., Rep M., Grivell L.A. Nucleic Acids Res. 20:4423-4428(1992).

477. RIO1/ZK632.3/MJ0444 family signature
The following uncharacterized proteins are evolutionary related [1]: - Yeast protein RIO1. - Caenorhabditis elegans hypothetical protein ZK632.3. - Methanococcus jannaschii hypothetical protein MJ0444. - Thermoplasma acidophilum hypothetical protein if rpoA2 3'region.The eukaryotic members of this family are proteins of about 55 to 60 Kd, while the archebacterial ones are half that size. The central part of these proteins is highly conserved. The best conserved region is used as a signature pattern.

Consensus pattern: [LIVM]-V-H-[GA]-D-L-S-E-[FY]-N-x-[LIVM]

[1] Bairoch A. Unpublished observations (1997).

478. (RIP) Shiga/ricin ribosomal inactivating toxins active site signature. A number of bacterial and plant toxins act by inhibiting protein synthesis in eukaryotic cells. The toxins of the Shiga and ricin family inactivate 60S ribosomal subunits by an N-glycosidic cleavage which releases a specific adenine base from the sugar-phosphate backbone of 28S rRNA [1,2,3]. The toxins which are known to function in this manner are: - Shiga toxin from Shigella dysenteriae [4]. This toxin is composed of one copy of an enzymatically active A subunit and five copies of a B subunit responsible for binding the toxin complex to specific

receptors on the target cell surface. - Shiga-like toxins (SLT) are a group of Escherichia coli toxins very similar in their structure and properties to Shiga toxin. The sequence of two types of these toxins, SLT-1 [5] and SLT-2 [6], is known. - Ricin, a potent toxin from castor bean seeds. Ricin consists of two glycosylated chains linked by a disulfide bond. The A chain is enzymatically active. The B chain is a lectin with a binding preference for galactosides. Both chains are encoded by a single polypeptidic precursor. Ricin is classified as a type-II ribosome-inactivating protein (RIP); other members of this family are agglutinin, also from castor bean, and abrin from the seeds of the bean Abrus precatorius [7]. - Single chain ribosome-inactivating proteins (type-I RIP) from plants. Examples of such proteins are: barley protein synthesis inhibitors I and II, mongolian snake-gourd trichosanthin, sponge gourd luffin-A and -B, garden four-o'clock MAP, common pokeberry PAP-S and soapwort saporin-6 [7].All these toxins are structurally related. A conserved glutamic residue has been implicated [8] in the catalytic mechanism; it is located near a conserved arginine which also plays a role in catalysis [9]. The signature that has been developed for these proteins includes these catalytic residues.

Consensus pattern: [LIVMA]-x-[LIVMSTA](2)-x-E-[SAGV]-[STAL]-R-[FY]-[RKNQS]-x-[LIVM]-[EQS]-x(2)-[LIVMF] [E and R are active site residues]-

[1] Endo Y., Tsurugi K., Takeda Y., Ogasawara T., Igarashi K. Eur. J. Biochem. 171:45-50(1988).[2] May M.J., Hartley M.R., Roberts L.M., Krieg P.A., Osborn R.W., Lord J.M. EMBO J. 8:301-308(1989).[3] Funatsu G., Islam M.R., Minami Y., Sung-Sil K., Kimura M. Biochimie 73:1157-1161(1991).[4] Strockbine N.A., Jackson M.P., Sung L.M., Holmes R.K., O'Brien A.D. J. Bacteriol. 170:1116-1122(1988).[5] Calderwood S.B., Auclair F., Donohue-Rolfe A., Keusch G.T., Mekalanos J.J. Proc. Natl. Acad. Sci. U.S.A. 84:4364-4368(1987).[6] Jackson M.P., Neill R.J., O'Brien A.D., Holmes R.K., Newland J.W. FEMS Microbiol. Lett. 44:109-114(1987).[7] Barbieri L., Battelli M.G., Stirpe F. Biochim. Biophys. Acta 1154:237-282(1993).[8] Hovde C.J., Calderwood S.B., Mekalanos J.J., Collier R.J. Proc. Natl. Acad. Sci. U.S.A. 85:2568-2572(1988).[9] Monzingo A.F., Collins E.J., Ernst S.R., Irvin J.D., Robertus J.D. J. Mol. Biol. 233:705-715(1993).

479. Bacterial RNA polymerase, alpha chain (RNA pol A bac)

Members of this family include alpha subunit from eubacteria and alpha subunits from chloroplasts. The alpha subunit of RNA polymerase consists of two independently folded domains, referred to as amino-terminal and carboxyl terminal domains. The amino terminal domain is involved in the interaction with the other subunits of the RNA polymerase. The carboxyl-terminal domain interacts with the DNA and activators. The amino acid sequence of the alpha subunit is conserved in prokaryotic and chloroplast RNA polymerases. There are three regions of particularly strong conservation, two in the amino-terminal and one in the carboxyl-Comment: terminal [3].

[1] Zhang G, Darst SA; Science 1998;281:262-266. [2] Jeon YH, Negishi T, Shirakawa M, Yamazaki T, Fujita N, Ishihama A, Kyogoku Y; Science 1995;270:1495-1497. [3] Ebright RH, Busby S; Curr Opin Genet Dev 1995;5:197-203. [4] Murakami K, Kimura M, Owens JT, Meares CF, Ishihama A; Proc Natl Acad Sci USA 1997;94:1709-1714.

480. RNA polymerase beta subunit (RNA pol B)

RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases). Each RNA polymerase complex contains two related members of this family, in each case they are the two largest subunits.

[1] Falkenburg D, Dworniczak B, Faust DM, Bautz EK; J Mol Biol 1987;195:929-937.

481. RNA polymerases H / 23 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaebacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. Archaebacterial subunit H (gene rpoH) [1,2] is a small protein of about 8.5 to10 Kd, it is evolutionary related to the C-terminal part of a 23 Kd component shared by all three forms of eukaryotic RNA polymerases (gene RPB5 in yeast and POLR2E in mammals).As a signature pattern a conserved region was selected which is located at theN-terminal extremity of subunit H; this region contains two histidines that could play a role in the binding of a metal ion.

Consensus pattern: H-[NEI]-[LIVM]-V-P-x-H-x(2)-[LIVM]-x(2)-[DE]

[1] Klenk H.-P., Palm P., Lottspeich F., Zillig W. Proc. Natl. Acad. Sci. U.S.A. 89:407-410(1992).

[2] Thiru A., Hodach M., Eloranta J.J., Kostourou V., Weinzierl R.O., Matthews S.; J. Mol. Biol. 287:753-760(1999).

482. RNA polymerases K / 14 to 18 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNApolymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaebacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. A component of 14 to 18 Kd shared by all three forms of eukaryotic RNA polymerases and which has been sequenced in budding yeast (gene RPB6 orRPO26), in fission yeast (gene rpb6 or rpo15), in human and in African swine fever virus [1] is evolutionary related [2] to archaebacterial subunit K (gene rpoK). The archaebacterial protein is colinear with the C-terminal part of the eukaryotic subunit.

Consensus pattern: [ST]-x-[FY]-E-x-[AT]-R-x-[LIVM]-[GSA]-x-R-[SA]-x-Q

[1] Lu Z., Kutish G.F., Sussman M.D., Rock D.L. Nucleic Acids Res. 21:2940-2940(1993).

[2] McKune K., Woychik N.A. J. Bacteriol. 176:4754-4756(1994).

483. RNA polymerases L / 13 to 16 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNApolymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaebacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. It has been shown that small subunits of about 13 to 16 Kd found in all three types of eukaryotic polymerases are highly conserved. Subunits known to belong to this family are: - Budding yeast RPC19 subunit from RNA polymerases I and III [1]. - Budding yeast RPB11 subunit from RNA polymerase II [2]. - Mammalian RPB11 (gene POLR2K) from RNA

polymerase II. - Caenorhabditis elegans hypothetical protein F58A4.9. - Methanococcus jannaschii RNA polymerase subunit L (gene rpoL). - Sulfolobus acidocaldarius RNA polymerase subunit L (gene rpoL) [3].As a signature pattern a conserved region was selected which is located at the N-terminal extremity of these polymerase subunits; this region contains two cysteines that could play a role in the binding of a metal ion.

Consensus pattern: [DE](2)-H-[ST]-[LIVM]-[GAP]-N-x(11)-V-x-[FM]-x(2)-Y-x(3)- H-P

[1] Dequard-Chablat M., Riva M., Carles C., Sentenac A. J. Biol. Chem. 266:15300-15307(1991).
[2] Woychik N.A., McKune K., Lane W.S., Young R.A. Gene Expr. 3:77-82(1993).
[3] Langer D. EMBL/GenBank: X70805.

484. RNA polymerases N / 8 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaebacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. Archaebacterial subunit N (gene rpoN) [1] is a small protein of about 8 Kd, it is evolutionary related [2] to a 8.3 Kd component shared by all three forms of eukaryotic RNA polymerases (gene RPB10 in yeast and POLR2J in mammals) as well as to African swine fever virus protein CP80R [3].As a signature pattern a conserved region was selected which is located at the N-terminal extremity of these polymerase subunits; this region contains two cysteines that could play a role in the binding of a metal ion.

Consensus pattern: [LIVMF](2)-P-[LIVM]-x-C-F-[ST]-C-G-

[1] Langer D., Hain J., Thuriaux P., Zillig W. Proc. Natl. Acad. Sci. U.S.A. 92:5768-5772(1995).
[2] McKune K., Woychik N.A. J. Bacteriol. 176:4754-4756(1994).
[3] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).

485. Ribonuclease HII

[1] Mian IS; Nucleic Acids Res 1997;25:3187-3189.

486. Ribonuclease PH signature

Prokaryotic ribonuclease PH (EC 2.7.7.56) (RNase PH) [1] is a phosphorolyticexoribonuclease that removes nucleotide residues following the -CCA terminus of tRNA and adds nucleotides to the ends of RNA molecules by using nucleoside diphosphates as substrates. RNase PH is a conserved protein of about 240 amino-acid residues. It is evolutionary related to Caenorhabditis elegans hypothetical protein B0564.1.As a signature pattern, the most highly conserved region was selected which is located in the central part of these proteins.

Consensus sequence: C-[DE]-[LIVM](2)-Q-[GTA]-D-G-[SG]-x(2)-[TA]-A

[1] Kelly K.O., Deutscher M.P. J. Biol. Chem. 267:17153-17158(1992).

487. RanBP1 domain

[1] Di Matteo G, Fuschi P, Zerfass K, Moretti S, Ricordy R, Cenciarelli C, Tripodi M, Jansen-Durr P, Lavia P; Cell Growth Differ 1995;6:1213-1224.

488. Rhodanese signatures

Rhodanese (thiosulfate sulfurtransferase) (EC 2.8.1.1) [1,2] is an enzyme which catalyzes the transfer of the sulfane atom of thiosulfate to cyanide, to form sulfite and thiocyanate. In vertebrates, rhodanese is a mitochondrial enzyme of about 300 amino-acid residues involved in forming iron-sulfur complexes and cyanide detoxification. A cysteine residue takes part in the catalytic mechanism. Some bacterial proteins closely related to rhodanese are also thought to express a sulfotransferase activity. These are: - Azotobacter vinelandii rhdA. - Escherichia coli sseA [3]. - Saccharopolyspora erythraea cysA [4]. - Synechococcus strain PCC 7942 rhdA [5]. RhdA is a periplasmic protein probably involved in the transport of sulfur compounds. Two patterns for the rhodanese family were developed. They are based on highly conserved regions, one which is located in the N-terminal region, the other at the C-terminal extremity of the enzyme.

Consensus pattern: [FY]-x(3)-H-[LIV]-P-G-A-x(2)-[LIVF]
Consensus pattern: [FY]-[DEAP]-G-[SA]-W-x-E-[FYW]

[1] Westley J. Meth. Enzymol. 77:285-291(1981).
[2] Weiland K.L., Dooley T.P. Biochem. J. 275:227-231(1991).
[3] Rudd K.E. Unpublished observations (1993).
[4] Donadio S., Shafiee A., Hutchinson C.R. J. Bacteriol. 172:350-360(1990).
[5] Laudenbach D.E., Ehrhardt D., Green L., Grossman A.R. J. Bacteriol. 173:2751-2760(1991).

489. Ribonuclease III family signature

Prokaryotic ribonuclease III (EC 3.1.26.3) (gene rnc) [1] is an enzyme that digests double-stranded RNA. It is involved in the processing of ribosomal RNA precursors and of some mRNAs. RNase III is evolutionary related [2] to the following proteins: - Fission yeast pac1, a ribonuclease that probably inhibits mating and meiosis by degrading a specific mRNA required for sexual development. - Yeast ribonuclease III (gene RNT1), a dsRNA-specific nuclease that cleaves eukaryotic preribosomal RNA at various sites. - Caenorhabditis elegans hypothetical protein F26E4.13. - Paramecium bursaria chlorella virus 1 protein A464R. - Synechocystis strain PCC 6803 hypothetical protein slr0346. - Fission yeast hypothetical protein SpAC8A4.08c, a protein with a N-terminal helicase domain and a C-terminal RNase III domain. - Caenorhabditis elegans hypothetical protein K12H4.8, a protein with the same structure as SpAC8A4.08c.These proteins share regions of sequence similarity; one of which is a highly conserved stretch of 9 residues which has been developed as a signature pattern.

Consensus pattern: [DEQ]-[RQ]-[LM]-E-[FYW]-[LV]-G-D-[SAR]-

[1] Nashimoto H., Uchida H. Mol. Gen. Genet. 201:25-29(1985).
[2] Mian I.S. Nucleic Acids Res. 25:3187-3195(1997).

490. Rieske iron-sulfur protein signatures

Ubiquinol-cytochrome c reductase (EC 1.10.2.2) (also known as the bc1 complexor complex III) is one of the electron transport chains of mitochondria and of some aerobic prokaryotes; it catalyzes the oxidoreduction of ubiquinol and cytochrome c. In the chloroplast of plants and in cyanobacteria plastoquinone-plastocyanin reductase (EC 1.10.99.1) (also known as the b6f complex) is functionally similar and catalyzes the oxidoreduction of plastoquinol and cytochrome f. One of the components of these electron transfer systems is an iron-sulfur protein with a 2Fe-2S cluster, which is called the Rieske protein [1,2]. The Rieske protein contains approximately 190 amino acid residues. The iron-sulfur cluster is complexed to the protein through cysteine and histidine residues. Two perfectly conserved regions in Rieske proteins contains all the residuesthat bind the iron-sulfur cluster. Both regions contain two cysteines and a histidine. The first cysteine and the histidine are 2Fe-2S ligands while the remaining cysteines form a disulfide bond [3]. Two conserved regions were selected as signature patterns.

Consensus pattern: C-[TK]-H-L-G-C-[LIVST] [The first C and the H are 2Fe-2S ligands] [The second C is involved in a disulfide bond]
Consensus pattern: C-P-C-H-x-[GSA] [The first C and the H are 2Fe-2S ligands] [The second C is involved in a disulfide bond]

[1] Gatti F.L., Meinhardt S.W., Ohnishi T., Tzagoloff A. J. Mol. Biol. 205:421-435(1989).
[2] Kallas T., Spiller S., Malkin R. Proc. Natl. Acad. Sci. U.S.A. 85:5794-5798(1988).
[3] Iwata S., Saynovits M., Link T.A., Michel H. Structure 4:567-579(1996).

491. Ribosomal protein L1 signature

Ribosomal protein L1 is the largest protein from the large ribosomal subunit.In Escherichia coli, L1 is known to bind to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1, 2], groups: - Eubacterial L1. - Algal and plant chloroplast L1. - Cyanelle L1. - Archaebacterial L1. - Vertebrate L10A. - Yeast SSM1.As a signature pattern, the best conserved region was selected located in the central section of these proteins. It is located at the end of an alpha helix thought to be involved in RNA-binding.

Consensus pattern: [IM]-x(2)-[LIVA]-x(2,3)-[LIVM]-G-x(2)-[LMS]-[GSNH]-[PTKR]-[KRAV]-G-x-[LIMF]-P-[DENSTKQ]

[1] Nikonov S.V., Nevskaya N., Eliseikina I.A., Fomenkova N.P., Nikulin A., Ossina N., Garber M., Jonsson B.-H., Briand C., Al-Karadaghi S., Svensson L.A., Aevarsson A., Liljas A. EMBO J. 15:1350-1359(1996).

[2] Olvera J., Wool I.G. 2.3.CO;2-"Biochem. Biophys. Res. Commun. 220:954-957(1996).

492. Ribosomal protein L10 signature

Ribosomal protein L10 is one of the proteins from the large ribosomal subunit. L10 is a protein of 162 to 185 amino-acid residues which has only been found so far in eubacteria. A conserved region located in the N-terminal section of these proteins was used as a signature pattern.

Consensus pattern: [DEH]-x(2)-[GS]-[LIVMF]-[STN]-[VA]-x-[DEQK]-[LIVMA]-x(2)-[LIM]-R

493. Ribosomal protein L10e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Vertebrate L10 (QM) [1]. - Plant L10. - Caenorhabditis elegans L10 (F10B5.1). - Yeast L10 (QSR1). - Methanococcus jannaschii MJ0543.These proteins have 174 to 232 amino-acid residues. A conserved region located in the central section was selected as a signature pattern.

Consensus pattern: R-x-A-[FYW]-G-K-[PA]-x-G-x(2)-A-R-V

[1] Chan Y.-L., Diaz J.-J., Denoroy L., Madjar J.-J., Wool I.G. 2.3.CO;2-"Biochem. Biophys. Res. Commun. 255:952-956(1996).

494. Ribosomal protein L11 signature

Ribosomal protein L11 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L11 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups:

- Eubacterial L11.
- Plant chloroplast L11 (nuclear-encoded).
- Read algal chloroplast L11.
- Cyanelle L11.
- Archaebacterial L11.
- Mammalian L12.
- Plants L12.
- Yeast L12 (YL15).

L11 is a protein of 140 to 165 amino-acid residues. A conserved region located in the C-terminal section of these proteins was selected as a signature pattern. In Escherichia coli, the C-terminal half of L11 has been shown [3] to be in an extended and loosely folded conformation and is likely to be buried within the ribosomal structure.

Consensus pattern: [RKN]-x-[LIVM]-x-G-[ST]-x(2)-[SNQ]-[LIVM]-G-x(2)-[LIVM]-x(0,1)-[DENG]

[1] Pucciarelli G., Remacha M., Ballesta J.P.G.; Nucleic Acids Res. 18:4409-4416(1990).
[2] Otaka E., Hashimoto T., Mizuta K., Suzuki K.; Protein Seq. Data Anal. 5:301-313(1993).
[3] Choli T. Biochem. Int. 19:1323-1338(1989).

495. Ribosomal protein L7/L12 C-terminal domain

[1] Leijonmarck M, Liljas A; J Mol Biol 1987;195:555-579.

496. Ribosomal protein L13 signature

Ribosomal protein L13 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L13 is known to be one of the early assembly proteins of

the 50S ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L13.

- Plant chloroplast L13 (nuclear-encoded). - Red algal chloroplast L13.
- Archaebacterial L13. - Mammalian L13a (Tum P198). - Yeast Rp22 and Rp23.

L11 is a protein of 140 to 250 amino-acid residues. As a signature pattern, a conserved region was selected located in the C-terminal section of these proteins.

Consensus pattern: [LIVM]-[KRV]-[GK]-M-[LIV]-[PS]-x(4,5)-[GS]-[NQEKRA]-x(5)-[LIVM]-x-[AIV]-[LFY]-x-[GDN]

[1] Chan Y.-L., Olvera J., Glueck A., Wool I.G. J. Biol. Chem. 269:5589-5594(1994).

497. Ribosomal protein L13e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Vertebrate L13 (was previously known as Breast Basic Conserved protein 1 (BBC1)). - Drosophila L13. - Plant L13. - Yeast probable L13 (YM9375.11c).

These proteins have 199 to 218 amino-acid residues. As a signature pattern, a stretch of about 16 residues in the first third of these proteins selected.

-Consensus pattern: [KR]-Y-x(2)-K-[LIVM]-R-[STA]-G-[KR]-G-F-[ST]-L-x-E

[1] Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 201:102-107(1994).

498. Ribosomal protein L14 signature

Ribosomal protein L14 is one of the proteins from the large ribosomal subunit. In eubacteria, L14 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L14. - Algal and plant chloroplast L14. - Cyanelle L14.

- Archaebacterial L14. - Yeast L17A. - Mammalian L23.
- Caenorhabditis elegans L23 (B0336.10). - Higher eukaryotes mitochondrial L14.

- Yeast mitochondrial Yml38 (gene MRPL38).

L14 is a protein of 119 to 137 amino-acid residues. As a signature pattern, a conserved region located in the C-terminal half of these proteins was selected.

-Consensus pattern: [GA]-[LIV](3)-x(9,10)-[DNS]-G-x(4)-[FY]-x(2)-[NT]-x(2)-V-[LIV]

[1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

499. Ribosomal protein L15 signature

Ribosomal protein L15 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L15 is known to bind the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L15. - Plant chloroplast L15 (nuclear-encoded).

- Archaebacterial L15. - Vertebrate L27a. - Tetrahymena thermophila L29.
- Fungi L27a (L29, CRP-1, CYH2).

L15 is a protein of 144 to 154 amino-acid residues. As a signature pattern, a conserved region was selected in the C-terminal section of these proteins.

-Consensus pattern: K-[LIVM](2)-[GASL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-[LIVMFCA]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G

[1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

500. Ribosomal protein L15e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian L15. - Insect L15. - Plant L15. - Yeast YL10 (L13) (Rp15r).
- Thermoplasma acidophilum L15.

These proteins have about 200 amino acid residues. As a signature pattern, a conserved region was selected located in the central section.

-Consensus pattern: [DE]-[KR]-A-R-x-L-G-[FY]-x-[SAP]-x(2)-G-[LIVMFY](4)-R-x-R-[IV]-x-R-G

[1] Zwickl P., Lupas A., Baumeister W.
Biochem. Biophys. Res. Commun. 209:684-688(1995).

501. Ribosomal protein L17 signature

Ribosomal protein L17 is one of the proteins from the large ribosomal subunit. L17 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L17.

- Yeast mitochondrial YmL8 (gene MRPL8).

Eubacterial L17 is a protein of 120 to 130 amino-acid residues. Yeast YmL8 is twice larger (238 residues), the sequence of its N-terminal half is colinear with that of eubacterial L17. As a signature pattern, a conserved region in the N-terminal section was selected.

-Consensus pattern: I-x-[ST]-[GT]-x(2)-[KR]-x-K-x(6)-[DE]-x-[LIMV]-[LIVMT]-T-x-[STAG]-[KR]

502. Ribosomal protein L18e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate L18 (known as L14 in Xenopus) [1]. - Plant L18.
- Yeast L18 (Rp28). - Halobacterium marismortui Hl29.
- Sulfolobus acidocaldarius Hl29e.

These proteins have 115 to 187 amino-acid residues., A stretch of about 13 residues in the first third of these proteins has been selected as a signature pattern.

-Consensus pattern: [KRE]-x-L-x(2)-[PS]-[KR]-x(2)-[RH]-[PSA]-x-[LIVM]-[NS]-[LIVM]-x-[RK]-[LIVM]

[1] Puder M., Barnard G.F., Staniunas R.J., Steele G.D. Jr., Chen L.B.
Biochim. Biophys. Acta 1216:134-136(1993).

503. Ribosomal L18p family

It has been shown that the amino terminal 93 amino acids of Swiss:P09895 are necessary and sufficient to bind 5S rRNA in vitro. The carboxyl-terminal half of the protein, comprising amino acids 151-296, serves to localize the protein to the nucleolus [1].

Number of members: 26

[1]
Medline: 96212235
Distinct domains in ribosomal protein L5 mediate 5 S rRNA binding and nucleolar localization.
Michael WM, Dreyfuss G;
J Biol Chem 1996;271:11571-11574.

504. Ribosomal protein L19 signature

Ribosomal protein L19 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L19 is known to be located at the 30S-50S ribosomal subunit interface and may play a role in the structure and function of the aminoacyl-tRNA binding site. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L19. - Red algal chloroplast L19. - Cyanelle L19.

L19 is a protein of 120 to 130 amino-acid residues.,

A conserved region in the C-terminal section has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x-[KRGTI]-x-[GSAI]-[KRQDA]-[VG]-[RSN]-X(0,1)-[KR]-[SA]-[KY]-[KLI]-[LYS]-Y-[LIM]-R

505. Ribosomal protein L19e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian ribosomal protein L19 [1]. - Drosophila ribosomal protein L19 [2].
- Slime mold (D. discoideum) vegetative specific protein V14 [3].
- Yeast ribosomal protein L19 (YL14). - Archebacterial ribosomal protein L19E.

These proteins have 148 to 203 amino-acid residues.

A stretch of about 20 residues in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: Q-[KR]-R-[LIVM]-x-[SA]-x(4)-[CV]-G-x(3)-[IV]-[WK]-[LIVF]-[DN]-P

[1] Chan Y.-L., Lin A., McNally J., Peleg D., Meyuhas O., Wool I.G. J. Biol. Chem. 262:1111-1115(1987).[2] Hart K., Klein T., Wilcox M. Mech. Dev. 43:101-110(1993).[3] Singleton C.K., Manning S.S., Ken R. Nucleic Acids Res. 17:9679-9692(1989).

506. Ribosomal protein L1e signature (Ribosomal_L4)

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists [1,2,3, 4] of: - Vertebrate L1 (L4). - Drosophila L1. - Plant L1. - Yeast L2 (Rp2). - Fission yeast L2. - Halobacterium marismortui HmaL4 (HL6). - Methanococcus jannaschii MJ0177.

These proteins have 246 (archaebacteria) to 427 (human) amino acids. A conserved region in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: N-x(3)-[KRM]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-x(7)-[RK]-[GS]-H

[1] Rafti F., Gargiulo G., Manzi A., Malva C., Graziani F. Nucleic Acids Res. 17:456-456(1989).[2] Presutti C., Villa T., Bozzoni I. Nucleic Acids Res. 21:3900-3900(1993).

[3] Bagni C., Mariottini P., Annesi F., Amaldi F. Biochim. Biophys. Acta 1216:475-478(1993).

[3] Arndt E., Kroemer W., Hatakeyama T. J. Biol. Chem. 265:3034-3039(1990).

507. Ribosomal protein L2 signature

Ribosomal protein L2 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L2 is known to bind to the 23S rRNA and to have peptidyltransferase activity. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial L2. - Algal and plant chloroplast L2. - Cyanelle L2. - Archaebacterial L2.

- Plant L2. - Slime mold L2. - Marchantia polymorpha mitochondrial L2.
- Paramecium tetraurelia mitochondrial L2. - Fission yeast K5, K37 and KD4.
- Yeast YL6. - Vertebrate L8.

The best conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: P-x(2)-R-G-[STAIV](2)-x-N-[APK]-x-[DE]

[1] Marty I., Meyer Y.
Nucleic Acids Res. 20:1517-1522(1992).
[2] Otaka E., Hashimoto T., Mizuta K., Suzuki K.
Protein Seq. Data Anal. 5:301-313(1993).

508. Ribosomal protein L20 signature

Ribosomal protein L20 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L20 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L20. - Algal and plant chloroplast L20.
- Cyanelle L20.

L20 is a protein of about 120 amino-acid residues. A conserved region located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: K-x(3)-[KRC]-x-[LIVM]-W-[IV]-[STNALV]-R-[LIVM]-[NS]-x(3)-[RKHS]

[1] Otaka E., Hashimoto T., Mizuta K., Suzuki K.
Protein Seq. Data Anal. 5:301-313(1993).

509. Ribosomal protein L21e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L21 [1]. - Entamoeba histolytica L21 [2].
- Caenorhabditis elegans L21 (C14B9.7). - Yeast L21E (URP1) [3].
- Halobacterium marismortui HL31 [4].

These proteins have 160 (eukaryotes) or 95 (archebacteria) amino-acid

residues. A conserved region in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: G-[DE]-x-V-x(10)-[GV]-x(2)-[FYH]-x(2)-[FY]-x-G-x-T-G

[1] Devi K.R.G., Chan Y.-L., Wool I.G. Biochem. Biophys. Res. Commun. 162:364-370(1989).

[2] Petter R., Rozenblatt S., Nuchamowitz Y., Mirelman D. Mol. Biochem. Parasitol. 56:329-333(1992).

[3] Jank B., Waldherr M., Schweyen R.J. Curr. Genet. 23:15-18(1993).

[4] Hatakeyama T., Kimura M. Eur. J. Biochem. 172:703-711(1988).

510. Ribosomal protein L21 signature

Ribosomal protein L21 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L21 is known to bind to the 23S rRNA in the presence of L20. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L21.

- Marchantia polymorpha chloroplast L21. - Cyanelle L21.
- Spinach chloroplast L21 (nuclear-encoded).

Eubacterial L21 is a protein of about 100 amino-acid residues, the mature form of the spinach chloroplast L21 has 200 residues. A conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-[RQ]-x(2)-[ST]

511. Ribosomal protein L22 signature

Ribosomal protein L22 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L22 is known to bind 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial L22.

- Algal and plant chloroplast L22 (in legumes L22 is encoded in the nucleus instead of the chloroplast). - Cyanelle L22. - Archaebacterial L22.
- Mammalian L17. - Plant L17. - Yeast YL17.

A conserved region located in the C- terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [RKQN]-x(4)-[RH]-[GAS]-x-G-[KRQS]-x(9)-[HDN]-[LIVM]-x-[LIVMS]-x-[LIVM]

[1] Gantt J.S., Baldauf S.L., Calie P.J., Weeden N.F., Palmer J.D. EMBO J. 10:3073-3078(1991).[2] Madsen L.H., Kreiberg J.D., Gausing K. Curr. Genet. 19:417-422(1991).

[3] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

512. Ribosomal protein L23 signature

Ribosomal protein L23 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L23 is known to bind a specific region on the 23S rRNA; in yeast, the corresponding protein binds to a homologous site on the 26S rRNA [1]. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [2,3,4], groups: - Eubacterial L23.

- Algal and plant chloroplast L23. - Archaebacterial L23. - Mammalian L23A.
- Caenorhabditis elegans L23A (F55D10.2). - Fungi L25.
- Yeast mitochondrial YmL41 (gene MRPL41 or MRP20).

A small conserved region in the C-terminal section of these proteins, which is probably involved in rRNA-binding has been selected as a signature pattern [2].

-Consensus pattern: [RK](2)-[AM]-[IVFYT]-[IV]-[RKT]-L-[STANEQK]-x(7)-[LIVMFT]

[1] El Baradi T.T.A.L., Raue H.A., van de Regt C.H.F., Verbree E.C., Planta R.J. EMBO J. 4:210-2107(1985).

[2] Raue H.A., Otaka E., Suzuki K. J. Mol. Evol. 28:418-426(1989).

[3] Fearon K., Mason T.L. J. Biol. Chem. 267:5162-5170(1992).

[4] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

513. Ribosomal protein L24 signature

Ribosomal protein L24 is one of the proteins from the large ribosomal subunit. L24 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L24.

- Plant chloroplast L24 (nuclear-encoded). - Red algal L24. - Vertebrate L26.
- Yeast L26 (YL33). - Archaebacterial HmaL24 (HL15).
- A probable ribosomal protein from Sulfolobus acidocaldarius [1].

In their mature form, these proteins have 103 to 150 amino-acid residues. A conserved stretch of 20 residues in their N-terminal section has been selected as a signature pattern.

-Consensus pattern: [GDEN]-D-x-V-x-[IV]-[LIVMA]-x-G-x(2)-[KRA]-[GNQ]-x(2,3)-[GA]-x-[IV]

[1] Ouzounis C., Kyrpides N., Sander C.
Nucleic Acids Res. 23:565-570(1995).

514. Ribosomal protein L24e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists [1] of:

- Mammalian ribosomal protein L24.
- Yeast ribosomal protein L30A/B (Rp29) (YL21).
- Kluyveromyces lactis ribosomal protein L30.
- Arabidopsis thaliana ribosomal protein L24 homolog.
- Haloarcula marismortui ribosomal protein HL21/HL22.
- Methanococcus jannaschii MJ1201.

These proteins have 60 to 160 amino-acid residues. The most conserved region, which is located in the N-terminal region of these proteins has been selected as a signature pattern.

-Consensus pattern: [FY]-x-[GSH]-x(2)-[IV]-x-P-G-x-G-x(2)-[FYV]-x-[KRHE]-x-D

[1] Chan Y.-L., Olvera J., Wool I.G.
Biochem. Biophys. Res. Commun. 202:1176-1180(1994).

515. Ribosomal protein L27 signature

Ribosomal protein L27 is one of the proteins from the large ribosomal subunit. L27 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial L27.

- Plant chloroplast L27 (nuclear-encoded). - Algal chloroplast L27.
- Yeast mitochondrial YmL2 (gene MRPL2 or MRP7).

The schematic relationship between these groups of proteins is shown below.

Eub. L27 NxxxxxxxxxAlgal L27 Nxxxxxxxxx

Plant L27 tttttNxxxxxxxxxxxxx

Yeast MRP7 tttNxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

***'t': transit peptide.

'N': N-terminal of mature protein.'*': position of the pattern.

-Consensus pattern: G-x-[LIVM](2)-x-R-Q-R-G-x(5)-G

[1] Elhag G.A., Bourque D.P. Biochemistry 31:6856-6864(1992).

[2] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

516. Ribosomal L28 family

The ribosomal 28 family includes L28 proteins from bacteria and chloroplasts. The L24 protein from yeast Swiss:P36525 also contains a region of similarity to prokaryotic L28 proteins. L24 from yeast is also found in the large ribosomal subunit

Number of members: 24

517. Ribosomal protein L29 signature

Ribosomal protein L29 is one of the proteins from the large ribosomal subunit. L29 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L29. - Red algal L29.

- Archaebacterial L29. - Mammalian L35 - Caenorhabditis elegans L35 (ZK652.4).
- Yeast L35.

L29 is a protein of 63 to 138 amino-acid residues.

A conserved region located in the central section of L29 has been selected as a signature pattern.

-Consensus pattern: [KNQS]-[PSTL]-x(2)-[LIMFA]-[KRGSAN]-x-[LIVYSTA]-[KR]-[KRHQS]-[DESTANRL]-[LIV]-A-[KRCQVT]-[LIVMA]

[1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

518. Ribosomal protein L3 signature

Ribosomal protein L3 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L3 is known to bind to the 23S rRNA and may participate in the formation of the peptidyltransferase center of the ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3,4], groups: - Eubacterial L3. - Red algal L3. - Cyanelle L3.

- Archaebacterial Halobacterium marismortui HmaL3 (HL1).
- Yeast L3 (also known as trichodermin resistance protein) (gene TCM1).
- Arabidopsis thaliana L3 (genes ARP1 and ARP2). - Mammalian L3 (L4).
- Mammalian mitochondrial L3. - Yeast mitochondrial YmL9 (gene MRPL9).

A conserved region located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: [FL]-x(6)-[DN]-x(2)-[AGS]-x-[ST]-x-G-[KRH]-G-x(2)-G-x(3)-R

[1] Arndt E., Kroemer W., Hatakeyama T. J. Biol. Chem. 265:3034-3039(1990).

[2] Graack H.-R., Grohmann L., Kitakawa M., Schaefer K.L., Kruft V. Eur. J. Biochem. 206:373-380(1992).

[3] Herwig S., Kruft V., Wittmann-Liebold B. Eur. J. Biochem. 207:877-885(1992).

[4] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

519. Ribosomal protein L30 signature

Ribosomal protein L30 is one of the proteins from the large ribosomal subunit. L30 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L30. - Archaebacterial L30.

- Drosophila L7. - Slime mold L7. - Mammalian L7. - Fungi L7 (YL8).
- Yeast mitochondrial L33.

L30 from eubacteria are small proteins of about 60 residues, those from archaebacteria are proteins of about 150 residues. Eukaryotic L7 are proteins of about 250 to 270 residues. The schematic relationship between the three groups of proteins is shown below.Eub. L30 NxxxxxxxxxxC

Arc. L30 NxxxxxxxxxxxxxxxxxxxxxxxxxxxxC

Euk. L7 NxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxC

'*******'*': position of the pattern.

The signature pattern for this family of ribosomal proteins spans the N-terminal half of the region common to all these proteins.

-Consensus pattern: [IVT]-[LIVM]-x(2)-[LF]-x-[LI]-x-[KRHQEG]-x(2)-[STNQH]-x-[IVT]-x(10)-[LMS]-[LIV]-x(2)-[LIVA]-x(2)-[LMFY]-[IVT]

[1] Mizuta K., Hashimoto T., Otaka E.
Nucleic Acids Res. 20:1011-1016(1992).

520. Ribosomal protein L31 signature

Ribosomal protein L31 is one of the proteins from the large ribosomal subunit. L31 is a protein of 66 to 97 amino-acid residues which has only been found so far in eubacteria and in some algal chloroplasts.

A conserved region located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: H-P-F-[FY]-[TI]-x(9)-G-R-[AIV]-x-[KRQ]

521. Ribosomal protein L31e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L31 [1]. - Chlamydomonas reinhardtii L31. - Yeast L34.
- Halobacterium marismortui HL30 [2].

These proteins have 87 to 128 amino-acid residues.

A conserved region, located in the central section has been selected as a signature pattern.

-Consensus pattern: V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AKH]-x-W-x-[KR]-G

[1] Tanaka T., Kuwano Y., Kuzumaki T., Ishikawa K., Ogata K.
Eur. J. Biochem. 162:45-48(1987).[2] Bergmann U., Arndt E.
Biochim. Biophys. Acta 1050:56-60(1990).

522. Ribosomal protein L33 signature

Ribosomal protein L33 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L33 has been shown to be on the surface of 50S subunit. L33 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial L33.

- Algal and plant chloroplast L33. - Cyanelle L33.

L33 is a small protein of 49 to 66 amino-acid residues. A conserved region located in the central section of L33 has been selected as a signature pattern.

-Consensus pattern: Y-x-[ST]-x-[KR]-[NS]-x(4)-[PATQ]-x(1,2)-[LIVM]-[EA]-x(2)-K-[FY]-[CSD]

[1] Kruft V., Kapp U., Wittmann-Liebold B. Biochimie 73:855-860(1991).

[2] Sharp P.M. Gene 139:129-130(1994).

[3] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

523. Ribosomal protein L34 signature

Ribosomal protein L34 is one of the proteins from the large subunit of the prokaryotic ribosome. It is a small basic protein of 44 to 51 amino-acid residues [1]. L34 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L34.

- Red algal chloroplast L34. - Cyanelle L34.

A conserved region that corresponds to the N-terminal half of L34 has been selected as a signature pattern.

-Consensus pattern: K-[RG]-T-[FYWL]-[EQS]-x(5)-[KRHS]-x(4,5)-G-F-x(2)-R

[1] Old I.G., Margarita D., Saint Girons I. Nucleic Acids Res. 20:6097-6097(1992).

524. Ribosomal protein L34e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L34. - Mosquito L31 [1]. - Plant L34 [2].

- Yeast putative ribosomal protein YIL052c. - Methanococcus jannaschii MJ0655.

These proteins have 89 to 129 amino-acid residues.

A conserved region located in the N-terminal section of these proteins has been

selected as a signature pattern.

-Consensus pattern: Y-x-[ST]-x-S-[NY]-x(5)-[KR]-T-P-G

[1] Lan Q., Niu L.L., Fallon A.M.
Biochim. Biophys. Acta 1218:460-462(1994).

[2] Gao J., Kim S.R., Chung Y.Y., Lee J.M., An G.
Plant Mol. Biol. 25:761-770(1994).

525. Ribosomal protein L35Ae signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate L35A. - Caenorhabditis elegans L35A (F10E7.7).
- Yeast L37A/L37B (Rp47). - Pyrococcus woesei L35A homolog [1].

These proteins have 87 to 110 amino-acid residues.

A highly conserved stretch of 22 residues in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P

[1] Ouzounis C., Kyrpides N., Sander C.
Nucleic Acids Res. 23:565-570(1995).

526. Ribosomal protein L36 signature

Ribosomal protein L36 is the smallest protein from the large subunit of the prokaryotic ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L36. - Algal and plant chloroplast L36. - Cyanelle L36.L36 is a small basic and cysteine-rich protein of 37 amino-acid residues. As a signature pattern, a conserved region that corresponds to positions 11 to 36 in L36 and includes three conserved cysteine residues has been developed.

Consensus pattern: C-x(2)-C-x(2)-[LIVM]-x-R-x(3)-[LIVMN]-x-[LIVM]-x-C-x(3,4)- [KR]-H-x-Q-x-Q-

[1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

527. Ribosomal protein L36e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L36 [1].

- Drosophila L36 (M(1)1B). - Caenorhabditis elegans L36 (F37C12.4).

- Candida albicans L39. - Yeast YL39.

These proteins have 99 to 104 amino acids.

A conserved region in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR]

[1] Chan Y.-L., Paz V., Olvera J., Wool I.G.
Biochem. Biophys. Res. Commun. 192:849-853(1993).

528. Ribosomal protein L39e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L39 [1]. - Plants L39. - Yeast L46 [2]. - Archebacterial L39e [3].

These proteins are very basic. About 50 residues long, they are the smallest proteins of eukaryotic-type ribosomes. A conserved region in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-[NHY]-W-R-R

[1] Lin A., McNally J., Wool I.G. J. Biol. Chem. 259:487-490(1984).

[2] Leer R.J., van Raamsdonk-Duin M.M.C., Kraakman P., Mager W.H.,
Planta R.J. Nucleic Acids Res. 13:701-709(1985).

[3] Ramirez C., Louie K.A., Matheson A.T. FEBS Lett. 250:416-418(1989).

529. Ribosomal L40e family

Bovine L40 has been identified as a secondary RNA binding protein [1]. L40 is fused to a ubiquitin protein [2].

Number of members: 27

[1]

Medline: 88203200

RNA binding proteins of the large subunit of bovine mitochondrial ribosomes.

Piatyszek MA, Denslow ND, O'Brien TW;
Nucleic Acids Res 1988;16:2565-2583.
[2]Medline: 96011832
The carboxyl extensions of two rat ubiquitin fusion proteins are ribosomal proteins S27a and L40.
Chan YL, Suzuki K, Wool IG;
Biochem Biophys Res Commun 1995;215:682-690.

530. (Ribosomal L44) Ribosomal protein L44e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L44 [1]. - Trypanosoma brucei L44.
- Caenorhabditis elegans L44 (C09H10.2). - Fungal L44 (L41).
- Halobacterium marismortui LA [2].

These proteins have 92 to 105 amino-acid residues.

A conserved region located in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: K-x-[TV]-K-K-x(2)-L-[KR]-x(2)-C

[1] Gallagher M.J., Chan Y.-L., Lin A., Wool I.G. DNA 7:269-273(1988).
[2] Bergmann U., Wittmann-Liebold B.
Biochim. Biophys. Acta 1173:195-200(1993

531. Ribosomal protein L5 signature

Ribosomal protein L5 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L5 is known to be involved in binding 5S RNA to the large ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3,4], groups: - Eubacterial L5.

- Algal chloroplast L5. - Cyanelle L5. - Archaebacterial L5. - Mammalian L11.
- Tetrahymena thermophila L21. - Slime mold L5 (V18). - Yeast L16 (39A).
- Plants mitochondrial L5.

L5 is a protein of about 180 amino-acid residues.

A conserved region, located in the first third of these

proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x(2)-[LIVM]-[STAVC]-[GE]-[QV]-x(2)-[LIVMA]-x-[STC]-x-[STAG]-[KRH]-x-[STA]

[1] Hatakeyama T., Hatakeyama T. Biochim. Biophys. Acta 1039:343-347(1990).

[2] Rosendahl G., Andreasen P.H., Kristiansen K. Gene 98:161-167(1991).

[3] Yang D., Gunther I., Matheson A.T., Auer J., Spicker G., Boeck A. Biochimie 73:679-682(1991).

[4] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

532. ribosomal L5P family C-terminus

This region is found associated with Ribosomal_L5.

Number of members: 60

533. Ribosomal protein L6 signatures

Ribosomal protein L6 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L6 is known to bind directly to the 23S rRNA and is located at the aminoacyl-tRNA binding site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3,4], groups: - Eubacterial L6.

- Algal chloroplast L6.
- Cyanelle L6.
- Archaebacterial L6.
- Marchantia polymorpha mitochondrial L6.
- Yeast mitochondrial YmL6 (gene MRPL6).
- Mammalian L9.
- Drosophila L9.
- Plants L9.
- Yeast L9 (YL11).

While all the above proteins are evolutionary related it is very difficult to derive a pattern that will find them all. Two patterns were therefore created, the first to detect

eubacterial, cyanelle and mitochondrial L6, the second to detect archaebacterial L6 as well as eukaryotic L9.

-Consensus pattern: [PS]-[DENS]-x-Y-K-[GA]-K-G-[LIVM]

-Consensus pattern: Q-x(3)-[LIVM]-x(2)-[KR]-x(2)-R-x-F-x-D-G-[LIVM]-Y-[LIVM]-x(2)-[KR]

[1] Suzuki K., Olvera J., Wool I.G. Gene 93:297-300(1990).

[2] Schwank S., Harrer R., Schueller H.-J., Schweizer E. Curr. Genet. 24:136-140(1993).

[3] Golden B.L., Ramakrishnan V., White S.W. EMBO J. 12:4901-4908(1993).

[4] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

534. Ribosomal protein L6e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian ribosomal protein L6 (L6 was previously known as TAX-responsive enhancer element binding protein 107).
- Caenorhabditis elegans ribosomal protein L6 (R151.3).
- Yeast ribosomal protein YL16A/YL16B.
- Mesembryanthemum crystallinum ribosomal protein YL16-like.

These proteins have 175 (yeast) to 287 (mammalian) amino acids. A highly conserved region in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: N-x(2)-P-L-R-R-x(4)-[FY]-V-I-A-T-S-x-K

535. Ribosomal protein L7Ae signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate L7A (SURF3) [1]. - Plant L7A. - Yeast L7A (YL5) (Rp6).
- Yeast protein NHP2 [2]. - Yeast hypothetical protein YEL026w.
- Bacillus subtilis hypothetical protein ylxQ. - Halobacterium marismortui Hs6.
- Methanococcus jannaschii MJ1203.

These proteins have 100 to 265 amino-acid residues.

A conserved region located in the central section has been selected as a signature pattern.

-Consensus pattern: [CA]-x(4)-[IV]-P-[FY]-x(2)-[LIVM]-x-[GSQ]-[KRQ]-x(2)-L-G

[1] Colombo P., Yon J., Garson K., Fried M. Proc. Natl. Acad. Sci. U.S.A. 89:6358-6362(1992).

[2] Kolodrubetz D., Burgum A. Yeast 7:79-90(1991).

536. Ribosomal protein L9 signature

Ribosomal protein L9 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L9 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial L9. - Cyanobacterial L9.

- Plant chloroplast L9 (nuclear-encoded). - Red algal chloroplast L9.

A conserved region, located in the N-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: G-x(2)-[GN]-x(4)-V-x(2)-G-[FY]-x(2)-N-[FY]-L-x(5)-[GA]-x(3)-[STN]

[1] Hoffman D.W., Davies C., Gerchman S.E., Kycia J.H., Porter S.J., White S.W., Ramakrishnan V. EMBO J. 13:205-212(1994).

[2] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

537. Ribosomal protein S10 signature

Ribosomal protein S10 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S10 is known to be involved in binding tRNA to the ribosomes. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S10.

- Algal chloroplast S10. - Cyanelle S10. - Archaebacterial S10.

- Marchantia polymorpha and Prototheca wickerhamii mitochondrial S10.

- Arabidopsis thaliana mitochondrial S10 (nuclear encoded). - Vertebrate S20.

- Plant S20. - Yeast URP2.

S10 is a protein of about 100 amino-acid residues.

A conserved region located in the center of these proteins has been selected as a signature pattern.

-Consensus pattern: [AV]-x(3)-[GDNSR]-[LIVMSTA]-x(3)-G-P-[LIVM]-x-[LIVM]-P-T

[1] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

538. Ribosomal protein S11 signature

Ribosomal protein S11 [1] plays an essential role in selecting the correct tRNA in protein biosynthesis. It is located on the large lobe of the small ribosomal subunit. S11 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups [2]: - Eubacterial S11.

- Algal and plant chloroplast S11. - Cyanelle S11. - Archaebacterial S11.
- Marchantia polymorpha and Prototheca wickerhamii mitochondrial S11.
- Acanthamoeba castellanii mitochondrial S11. - Neurospora crassa S14 (crp-2).
- Yeast S14 (RP59 or CRY1).
- Mammalian, Drosophila, Trypanosoma, and plant S14.
- Caenorhabditis elegans S14 (F37C12.9).

One of the best conserved regions in these proteins was selected as a signature pattern.

-Consensus pattern: [LIVMF]-x-[GSTAC]-[LIVMF]-x(2)-[GSTAL]-x(0,1)-[GSN]-[LIVMF]-x-[LIVM]-x(4)-[DEN]-x-T-P-x-[PA]-[STCH]-[DN]

[1] Kimura M., Kimura J., Hatakeyama T. FEBS Lett. 240:15-20(1988).
[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

539. Ribosomal protein S12 signature

Ribosomal protein S12 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S12 is known to be involved in the translation initiation step. It is a very basic protein of 120 to 150 amino-acid residues. S12 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S12. - Archaebacterial S12.

- Algal and plant chloroplast S12. - Cyanelle S12.

- Protozoa and plant mitochondrial S12. - Yeast S28.
- Drosophila mitochondrial protein tko (Technical KnockOut). - Mammalian S23.

The best conserved regions in these proteins, located in the center of each sequence have been selected as a signature pattern.

-Consensus pattern: [RK]-x-P-N-S-[AR]-x-R

[1] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

540. Ribosomal protein S12e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Vertebrate S12 [1].
- Trypanosoma brucei S12 [2]. - Caenorhabditis elegans S12 (F54E7.2).
- Drosophila S12. - Yeast S12.

These proteins have 130 to 150 amino acids.

A conserved region in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: A-L-[KRQP]-x-V-L-x(2)-[SA]-x(3)-[DN]-G-L

[1] Lin A., Chan Y.-L., Jones R., Wool I.G.
J. Biol. Chem. 262:14343-14351(1987).[2] Marchal C., Ismaili N., Pays E.
Mol. Biochem. Parasitol. 57:331-334(1993).

541. Ribosomal protein S13 signature

Ribosomal protein S13 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S13 is known to be involved in binding fMet-tRNA and, hence, in the initiation of translation. It is a basic protein of 115 to 177 amino-acid residues and belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S13.
- Plant chloroplast S13 (nuclear encoded). - Red algal chloroplast S13.
- Cyanelle S13. - Archaebacterial S13. - Plant mitochondrial S13.
- Mammalian and plant S18.

The best conserved regions in these proteins, located in their C-terminal part have been selected as a signature pattern.

-Consensus pattern: [KRQS]-G-x-R-H-x(2)-[GSNH]-x(2)-[LIVMC]-R-G-Q

[1] Chan Y.-L., Paz V., Wool I.G.
Biochem. Biophys. Res. Commun. 178:1212-1218(1991).

[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

542. Ribosomal protein S14p/S29e (Ribosomal protein S14 signature)

Ribosomal protein S14 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S14 is known to be required for the assembly of 30S particles and may also be responsible for determining the conformation of 16S rRNA at the A site. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups:

- Eubacterial S14.
- Algal and plant chloroplast S14.
- Cyanelle S14.
- Archaebacterial Methanococcus vannielii S14.
- Plant mitochondrial S14.
- Yeast mitochondrial MRP2.
- Mammalian S29.
- Yeast YS29A/B.

S14 is a protein of 53 to 115 amino-acid residues. Our signature pattern is based on the few conserved positions located in the center of these proteins.

Consensus pattern: [RP]-x(0,1)-C-x(11,12)-[LIVMF]-x-[LIVMF]-[SC]-[RG]-x(3)-[RN]

[1] Chan Y.-L., Suzuki K., Olvera J., Wool I.G. Nucleic Acids Res. 21:649-655(1993).

[2] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

543. Ribosomal protein S15 signature

Ribosomal protein S15 is one of the proteins from the small ribosomal subunit. In Escherichia coli, this protein binds to 16S ribosomal RNA and functions at early steps in ribosome assembly. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S15.

- Archaebacterial Halobacterium marismortui HmaS15 (HS11).
- Plant chloroplast S15. - Yeast mitochondrial S28. - Mammalian S13.
- Brugia pahangi and Wuchereria bancrofti S13 (S15). - Yeast S13 (YS15).

S15 is a protein of 80 to 250 amino-acid residues.

A conserved region located in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x(2)-H-[LIVMFY]-x(5)-D-x(2)-[SAGN]-x(3)-[LF]-x(9)-[LIVM]-x(2)-[FY]

[1] Dang H., Ellis S.R.
Nucleic Acids Res. 18:6895-6901(1990).

[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

544. Ribosomal protein S16 signature

Ribosomal protein S16 is one of the proteins from the small ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial S16.
- Algal and plant chloroplast S16.
- Cyanelle S16.
- Neurospora crassa mitochondrial S24 (cyt-21).

S16 is a protein of about 100 amino-acid residues. A conserved region located in the N-terminal extremity of these proteins has been selected as a signature pattern.

Consensus pattern: [LIVMT]-x-[LIVM]-[KR]-L-[STAK]-R-x-G-[AKR]

[1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

545. Ribosomal protein S17 signature

Ribosomal protein S17 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S17 is known to bind specifically to the 5' end of 16S ribosomal RNA and is thought to be involved in the recognition of termination

codons. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial S17.

- Plant chloroplast S17 (nuclear encoded). - Red algal chloroplast S17.
- Cyanelle S17. - Archaebacterial S17. - Mammalian and plant cytoplasmic S11.
- Yeast S18a and S18b (RP41; YS12).

The best conserved regions located in the C-terminal sections of these proteins have been selected as a signature pattern.

-Consensus pattern: G-D-x-[LIV]-x-[LIVA]-x-[QEK]-x-[RK]-P-[LIV]-S

[1] Gantt J.S., Thompson M.D. J. Biol. Chem. 265:2763-2767(1990).

[2] Herfurth E., Hirano H., Wittmann-Liebold B. Biol. Chem. Hoppe-Seyler 372:955-961(1991).

[3] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

546. Ribosomal protein S17e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrates S17 [1]. - Drosophila S17 [2]. - Neurospora crassa S17 (crp-3).
- Yeast S17a (RP51A) and S17b (RP51B) [3]. - Methanococcus jannaschii MJ0245.

These proteins have from 63 (in archebacteria) to 130 to 146 amino acids and are highly conserved. A region in the central part of these proteins has been selected as a signature.

-Consensus pattern: A-x-I-x-[ST]-K-x-L-R-N-[KR]-I-A-G-[FY]-x-T-H

[1] Chen I.-T., Roufa D.J. Gene 70:107-116(1988).

[2] Maki C., Rhoads D.D., Stewart M.J., van Slyke B., Denell R.E., Roufa D.J. Gene 79:289-298(1989).[3] Abovich N., Rosbash M. Mol. Cell. Biol. 4:1871-1879(1984).

547. Ribosomal protein S18 signature

Ribosomal protein S18 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S18 has been involved in aminoacyl-tRNA binding[1]. It appears to be situated at the tRNA A-site of the ribosome. It belongs to a family of ribosomal proteins

which, on the basis of sequence similarities[2], groups: - Eubacterial S18. - Algal and plant chloroplast S18. - Cyanelle S18.As a signature pattern, a conserved region in the central section of the protein has been selected. This region contains two basic residues which may be involved in RNA-binding.-

Consensus pattern: [IV]-[DY]-Y-x(2)-[LIVMT]-x(2)-[LIVM]-x(2)-[FYT]-[LIVM]- [ST]-[DERP]-x-[GY]-K-[LIVM]-x(3)-R-[LIVMAS]-

[1] McDougall J., Choli T., Kruft V., Kapp U., Wittmann-Liebold B. FEBS Lett. 245:253-260(1989).[2] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

548. Ribosomal protein S19 signature

Ribosomal protein S19 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S19 is known to form a complex with S13 that binds strongly to 16S ribosomal RNA. S19 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S19.

- Algal and plant chloroplast S19. - Cyanelle S19. - Archaebacterial S19.
- Plant mitochondrial S19. - Eukaryotic S15 ('rig' protein).

S19 is a protein of 88 to 144 amino-acid residues. Our signature pattern is based on the few conserved positions located in the C-terminal section of these proteins.

-Consensus pattern: [STDNQ]-G-[KRQM]-x(6)-[LIVM]-x(4)-[LIVM]-[GSD]-x(2)-[LF]-[GAS]-[DE]-F-x(2)-[ST]

[1] Kitagawa M., Takasawa S., Kikuchi N., Itoh T., Teraoka H., Yamamoto H., Okamoto H. FEBS Lett. 283:210-214(1991).

[2] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

549. Ribosomal protein S19e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities [1,2]. One of these families consists of: - Mammalian S19. - Drosophila S19.

- Ascaris lumbricoides S19g (ALEP-1) and S19s. - Yeast YS16 (RP55A and RP55B).
- Aspergillus S16. - Halobacterium marismortui HS12.

These proteins have 143 to 155 amino acids.

A well conserved stretch of 20 residues in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: P-x(6)-[SAN]-x(2)-[LIVMA]-x-R-x-[ALIV]-[LV]-Q-x-L-[EQ]

[1] Etter A., Aboutanos M., Tobler H., Mueller F.
Proc. Natl. Acad. Sci. U.S.A. 88:1593-1596(1991).

[2] Suzuki K., Olvera J., Wool I.G. Biochimie 72:299-302(1990).

550. Ribosomal protein S2 signatures

Ribosomal protein S2 is one of the proteins from the small ribosomal subunit. S2 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S2. - Algal and plant chloroplast S2.

- Cyanelle S2. - Archaebacterial S2.
- Higher eukaryotes P40 (previously thought to be a laminin receptor).
- Yeast NAB1. - Plant mitochondrial S2. - Yeast mitochondrial MRP4.

S2 is a protein of 235 to 394 amino-acid residues.

Two conserved regions have been selected as signature patterns. One is located in the N-terminal section and the other in the central section.

-Consensus pattern: [LIVMFA]-x(2)-[LIVMFYC](2)-x-[STAC]-[GSTANQEKR]-[STALV]-[HY]-[LIVMF]-G

-Consensus pattern: P-x(2)-[LIVMF](2)-[LIVMS]-x-[GDN]-x(3)-[DENL]-x(3)-[LIVM]-x-E-x(4)-[GNQKRH]-[LIVM]-[AP]

[1] Davis S.C., Tzagoloff A., Ellis S.R.
J. Biol. Chem. 267:5508-5514(1992).

[2] Tohgo A., Takasawa S., Munakata H., Yonekura H., Hayashi N., Okamoto H.
FEBS Lett. 340:133-138(1994).

551. Ribosomal protein S21 signature

Ribosomal protein S21 is one of the proteins from the small ribosomal subunit. So far S21 has only been found in eubacteria. It is a protein of 55 to 70 amino-acid residues. A conserved region in the N-terminal section of the protein has been selected as a signature pattern.

Consensus pattern: [DE]-x-A-[LIY]-[KR]-R-F-K-[KR]-x(3)-[KR]

552. Ribosomal protein S21e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian S21 [1].

- Caenorhabditis elegans S21 (F37C12.11). - Rice S21 [2].

- Yeast S21 (Ys25) [3]. - Fission yeast S28 [4].

These proteins have 82 to 87 amino acids.

A perfectly conserved nonapeptide in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: L-Y-V-P-R-K-C-S-[SA]

[1] Bhat K.S., Morrison S.G. Nucleic Acids Res. 21:2939-2939(1993).

[2] Nishi R., Hashimoto H., Uchimiya H., Kato A. Biochim. Biophys. Acta 1216:113-114(1993).[3] Suzuki K., Otaka E. Nucleic Acids Res. 16:6223-6223(1988).[4] Itoh T., Okata E., Matsui K.A. Biochemistry 24:7418-7423(1985).

553. Ribosomal protein S24e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate S24 [1]. - Yeast Rp50. - Mucor racemosus S24 [2].

- Halobacterium marismortui HS15 [3]. - Methanococcus jannaschii MJ0394.

These proteins have 101 to 148 amino acids.

A well conserved stretch in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: [FYA]-G-x(2)-[KR]-[STA]-x-G-[FY]-[GA]-x-[LIVM]-Y-[DN]-[SDN]

[1] Brown S.J., Jewell A., Maki C.G., Roufa D.J. Gene 91:293-296(1990).

[2] Sosa L., Fonzi W.A., Sypherd P.S. Nucleic Acids Res. 17:9319-9331(1989).[3] Kimura J., Arndt E., Kimura M. FEBS Lett. 224:65-70(1987).

554. Ribosomal protein S26e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian S26 [1]. - Octopus S26 [2]. - Drosophila S26 (DS31) [3]. - Plant cytoplasmic S26. - Fungi S26 [4].

These proteins have 114 to 127 amino acids.

A conserved octapeptide in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: [YH]-C-V-S-C-A-I-H

[1] Kuwano Y., Nakanishi O., Nabeshima Y., Tanaka T., Ogata K. J. Biochem. 97:983-992(1985).[2] Zinov'eva R.D., Tomarev S.I. Dokl. Akad. Nauk SSSR 304:464-469(1989).

[3] Itoh N., Ohta K., Ohta M., Kawasaki T., Yamashina I. Nucleic Acids Res. 17:2121-2121(1989).[4] Wu M., Tan H. Gene 150:401-402(1994).

555. Ribosomal protein S28e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian S28 [1]. - Plant S28 [2]. - Fungi S33 [3]. - Methanococcus jannaschii MJ1202.

These proteins have from 64 to 78 amino acids.

A highly conserved nonapeptide from the C-terminal extremity of these proteins has been selected as a signature pattern.

-Consensus pattern: E-[ST]-E-R-E-A-R-x-L

[1] Chan Y.-L., Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 179:314-318(1991).

[2] Hwang I., Goodman H.M. Plant Physiol. 102:1357-1358(1993).

[3] Hoekstra R., Ferreira P.M., Bootsman T.C., Mager W.H., Planta R.J. Yeast 8:949-959(1992).

556. Ribosomal protein S3Ae signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S3A (was originally known as v-fos transformation effector protein). - Caenorhabditis elegans S3A (F56F3.5).
- Plant cytoplasmic S3A (CYC07) [1]. - Yeast Rp10 (PLC1 and PLC2).
- Fission yeast Rp10 (SpAC13G6.02c). - Methanococcus jannaschii MJ0980.

These proteins have from 220 to 250 amino acids.

A conserved stretch in their N-terminal section was selected as a signature pattern.

-Consensus pattern: [LIV]-x-[GH]-R-[IV]-x-E-x-[SC]-L-x-D-L

[1] Liu J.H., Reid D.M.
Plant Physiol. 109:338-338(1995).

557. Ribosomal protein S3 signature

Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S3.

- Algal and plant chloroplast S3. - Cyanelle S3. - Archaebacterial S3.
- Plant mitochondrial S3. - Vertebrate S3. - Insect S3.
- Caenorhabditis elegans S3 (C23G10.3). - Yeast S3 (Rp13).

S3 is a protein of 209 to 559 amino-acid residues.

A conserved region located in the C-terminal section has been selected as a signature pattern.

-Consensus pattern: [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-G

[1] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

558. Ribosomal protein S4 signature

Ribosomal protein S4 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S4 is known to bind directly to 16S ribosomal RNA. Mutations in S4 have been shown to increase translational error frequencies.

It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S4. - Algal and plant chloroplast S4.

- Cyanelle S4. - Archaebacterial S4. - Mammalian S9. - Yeast YS11 (SUP45).
- Marchantia polymorpha mitochondrial S4. - Dictyostelium discoideum rp1024.
- Yeast protein NAM9 [3]. NAM9 has been characterized as a suppressor for ochre mutations in mitochondrial DNA. It could be a ribosomal protein that acts as a suppressor by decreasing translation accuracy.

S4 is a protein of 171 to 205 amino-acid residues (except for NAM9 which is much larger). The signature pattern for this protein is based on a conserved region located in the central section of these proteins.

-Consensus pattern: [LIVM]-[DE]-x-R-[LI]-x(3)-[LIVMC]-[VMFYHQ]-[KRT]-x(3)-[STAGCVF]-x-[ST]-x(3)-[SAI]-[KR]-x-[LIVMF](2)

[1] Mizuta K., Hashimoto T., Suzuki K.I., Otaka E.
Nucleic Acids Res. 19:2603-2608(1991).

[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

[3] Boguta M., Dmochowska A., Borsuk P., Wrobel K., Gargouri A., Lazowska J., Slonimski P., Szczesniak B., Kruszewska A.
Mol. Cell. Biol. 12:402-412(1992).

559. Ribosomal protein S4e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S4 [1]. Two highly similar isoforms of this protein exist : one coded by a gene on chromosome Y, and the other on chromosome X.
- Plant cytoplasmic S4 [2] - Yeast S7 (YS6). - Archebacterial S4e.

These proteins have 233 to 264 amino acids.

A highly conserved stretch of 15 residues in their N-terminal section has been selected as a signature pattern. Four positions in this region are positively charged residues.

-Consensus pattern: H-x-K-R-[LIVMF]-[SANK]-x-P-x(2)-[WY]-x-[LIVM]-x-[KRP]

[1] Fisher E.M., Beer-Romero P., Brown L.G., Ridley A., McNeil J.A.,
Lawrence J.B., Willard H.F., Bieber F.R., Page D.C.

Cell 63:1205-1218(1990).
[2] Braun H.P., Emmermann M., Mentzel H., Schmitz U.K.
Biochim. Biophys. Acta 1218:435-438(1994).

560. Ribosomal protein S5 signature

Ribosomal protein S5 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S5 is known to be important in the assembly and function of the 30S ribosomal subunit. Mutations in S5 have been shown to increase translational error frequencies. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S5.
- Cyanelle S5. - Red algal chloroplast S5. - Archaebacterial S5.
- Mammalian S2 (LLrep3). - Caenorhabditis elegans S2 (C49H3.11).
- Drosophila S2. - Plant S2. - Yeast S4 (SUP44). - Fungi mitochondrial S5.

S5 is a protein of 166 to 254 amino-acid residues. The signature pattern for this protein is based on a conserved region, rich in glycine residues, and located in the N-terminal section of these proteins.

-Consensus pattern: G-[KRQ]-x(3)-[FY]-x-[ACV]-x(2)-[LIVMA]-[LIVM]-[AG]-[DN]-x(2)-G-x-[LIVM]-G-x-[SAG]-x(5,6)-[DEQ]-[LIVMA]-x(2)-A-[LIVMF]

[1] All-Robyn J.A., Brown N., Otaka E., Liebman S.W.
Mol. Cell. Biol. 10:6544-6553(1990).[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

561. Ribosomal protein S6 signature

Ribosomal protein S6 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S6 is known to bind together with S18 to 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial S6. - Red algal chloroplast S6.
- Cyanelle S6.

S6 is a protein of 95 to 208 amino-acid residues. The signature pattern for this protein is based on a conserved region located in the N-terminal section of these proteins.

-Consensus pattern: G-x-[KRC]-[DENQRH]-L-[SA]-Y-x-I-[KRNSA]

562. Ribosomal protein S6e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S6 [1]. - Drosophila S6 [2]. - Plant S6 [3]. - Yeast S10 (YS4).
- Halobacterium marismortui HS13 [4]. - Methanococcus jannaschii MJ1260.

S6 is the major substrate of protein kinases in eukaryotic ribosomes [5]; it may have an important role in controlling cell growth and proliferation through the selective translation of particular classes of mRNA.

These proteins have 135 to 249 amino acids.

A conserved stretch of 12 residues in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-[STAMR]-G-G-x-D-x(2)-G-x-P-M

[1] Franco R., Rosenfeld M.G. J. Biol. Chem. 265:4321-4325(1990).

[2] Watson K.L., Konrad K.D., Woods D.F., Bryant P.J. Proc. Natl. Acad. Sci. U.S.A. 89:11302-11306(1992).

[3] Hansen G., Estruch J.J., Spena A. Nucleic Acids Res. 20:5230-5230(1992).

[4] Kimura M., Arndt E., Hatakeyama T., Hatakeyama T., Kimura J. Can. J. Microbiol. 35:195-199(1989).

[5] Bandi H.R., Ferrari S., Krieg J., Meyer H.E., Thomas G. J. Biol. Chem. 268:4530-4533(1993).

563. Ribosomal protein S7 signature

Ribosomal protein S7 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S7 is known to bind directly to part of the 3'end of 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial S7.

- Algal and plant chloroplast S7. - Cyanelle S7. - Archaebacterial S7.
- Plant mitochondrial S7. - Mammalian S5. - Plant S5.
- Caenorhabditis elegans S5 (T05E11.1).

The best conserved region located in the N-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [DENSK]-x-[LIVMDET]-x(3)-[LIVMFTA](2)-x(6)-G-K-[KR]-x(5)-[LIVMF]-[LIVMFC]-x(2)-[STAC]

[1] Klussmann S., Franke P., Bergmann U., Kostka S., Wittmann-Liebold B. Biol. Chem. Hoppe-Seyler 374:305-312(1993).

[2] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

[3] Ignatovich O., Cooper M., Kulesza H.M., Beggs J.D. Nucleic Acids Res. 23:4616-4619(1995).

564. Ribosomal protein S7e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian S7.
- Xenopus S8.
- Insect S7.
- Yeast probable ribosomal protein S7 (N2212).
- Fission yeast probable ribosomal protein S7 (SpAC18G6.13c).

These proteins have about 200 amino acids. A highly conserved stretch of 14 residues which is located in the central section and which is rich in charged residues was selected as a signature pattern.

Consensus pattern: [KR]-L-x-R-E-L-E-K-K-F-[SAP]-x-[KR]-H

[1] Salazar C.E., Mills-Hamm D.M., Kumar V., Collins F.H. Nucleic Acids Res. 21:4147-4147(1993).

565. Ribosomal protein S8 signature

Ribosomal protein S8 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S8 is known to bind directly to 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of sequence

similarities [1], groups: - Eubacterial S8. - Algal and plant chloroplast S8.

- Cyanelle S8. - Archaebacterial S8. - Marchantia polymorpha mitochondrial S8.

- Mammalian S15A. - Plant S15A. - Yeast S22 (S24).

The best conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [GE]-x(2)-[LIV](2)-[STY]-[ST]-x(2)-G-[LIVM](2)-x(4)-[AG]-[KRHAYI]

[1] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

566. Ribosomal protein S8e signature

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian S8. - Caenorhabditis elegans S8 (F42C5.8). - Leishmania major S8.

- Plant S8. - Yeast S8 (S14) (Rp19). - Archebacterial S8e.

These proteins have either about 220 amino acids (in eukaryotes) or about 125 amino acids (in archebacteria). A conserved stretch which is located in the N-terminal section and which is rich in positively charged residues has been selected as a signature pattern.

-Consensus pattern: [KR]-x(2)-[ST]-G-[GA]-x(5)-[HR]-[KG]-[KR]-x-K-x-E-[LM]-G

[1] Engemann S., Herfurth E., Briesemeister U., Wittmann-Liebold B.
J. Protein Chem. 14:189-195(1995).

567. Ribosomal protein S9 signature

Ribosomal protein S9 is one of the proteins from the small ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S9. - Algal chloroplast S9.

- Cyanelle S9. - Archaebacterial S9. - Mammalian S16. - Plant S16.

- Yeast mitochondrial ribosomal S9.

A conserved region containing many charged residues and located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: G-G-G-x(2)-[GSA]-Q-x(2)-[SA]-x(3)-[GSA]-x-[GSTAV]-[KR]-

[GSAL]-[LIF]

[1] Chan Y.-L., Paz V., Olvera J., Wool I.G. FEBS Lett. 263:85-88(1990).

[2] Otaka E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

568. Ribulose-phosphate 3-epimerase family signatures

Ribulose-phosphate 3-epimerase (EC 5.1.3.1) (also known as pentose-5-phosphate 3-epimerase or PPE) is the enzyme that converts D-ribulose 5-phosphate into D-xylulose 5-phosphate in Calvin's reductive pentose phosphate cycle. In Alcaligenes eutrophus two copies of the gene coding for PPE are known [1], one is chromosomally encoded (cbbEC), the other one is on a plasmid (cbbeP). PPE has been found in a wide range of bacteria, archebacteria, fungi and plants. The sequence of PPE is highly related to:

- Escherichia coli D-allulose-6-phosphate 3-epimerase (gene alsE).
- Escherichia coli protein sgcE.
- Mycoplasma genitalium hypothetical protein MG112.

All these proteins have from 209 to 241 amino acid residues.

Two conserved regions which are located respectively in the N-terminal and in the central part of these proteins have been selected as signature patterns.

-Consensus pattern: [LIVMF]-H-[LIVMFY]-D-[LIVM]-x-D-x(1,2)-[FY]-[LIVM]-x-N-x-[STAV]

-Consensus pattern: [LIVMA]-x-[LIVM]-M-[ST]-[VS]-x-P-x(3)-G-Q-x-F-x(6)-[NK]-[LIVMC]

[1] Kusian B., Yoo J.G., Bednarski R., Bowien B.
J. Bacteriol. 174:7337-7344(1992).

569. (Ricin B lectin) Similarity to lectin domain of ricin beta-chain, 3 copies.

This family consists of a triplicated domain involved in cell agglutination in ricin.

570. (Rotamase) PpiC-type peptidyl-prolyl cis-trans isomerase signature

Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase or rotamase) is an enzyme that accelerates protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides [1]. Most characterized PPiases belong to two families, the cyclophilin-type (see <PDOC00154>) and the the FKBP-type (see <PDOC00426>). Recently a third family has been discovered [2,3]. So far, the only biochemically characterized member of this family is the Escherichia coli protein parvulin (gene ppiC), a small (92 residues) cytoplasmic enzyme that prefers amino acid residues with hydrophobic side chains like leucine and phenylalanine in the P1 position of the peptides substrates. PpiC is evolutionary related to a number of proteins that are also probably PPiases:

- Escherichia coli and Haemophilus influenzae ppiD. PpiD is a PPIase which contains a periplasmic ppiC-like domain anchored to the inner membrane and which seems to be involved in the folding of outer membrane proteins.
- Escherichia coli surA. SurA is a periplasmic protein that contains two ppiC-like domains.
- Nitrogen-assimilating bacteria protein nifM which is involved in the activation and stabilization of the iron-component (nifH) of nitrogenase.
- Bacillus subtilis protein prsA, a membrane-bound lipoprotein involved in protein export.
- Lactococcus and lactobacillus protease maturation protein prtM, a membrane-bound lipoprotein involved in the maturation of a secreted serine proteinase. - Yeast protein ESS1/PTF1 (processing/termination factor 1).
- Drosophila protein dodo (gene dod). - Mammalian protein PIN1,
- Campylobacter jejuni cell binding factor 2 (CBF2), a secreted antigen.
- Bacillus subtilis hypothetical protein yacD.
- Helicobacter pylori hypothetical protein HP0175.
- A hypothetical slime mold protein.

A conserved region that contains a serine which could play a role in the catalytic mechanism of these enzymes has been selected as a signature pattern.

-Consensus pattern: F-[GSADEI]-x-[LVAQ]-A-x(3)-[ST]-x(3,4)-[STQ]-x(3,5)-[GER]-G-x-[LIVM]-[GS]

[1] Fischer G., Schmid F.X.

Biochemistry 29:2205-2212(1990).
[2] Rudd K.E., Sofia H.J., Koonin E.V., Plunkett G. III, Lazar S., Rouviere P.E. Trends Biochem. Sci. 20:14-15(1995).
[3] Rahfeld J.-U., Ruecknagel K.P., Schelbert B., Ludwig B., Hacker J., Mann K., Fischer G. FEBS Lett. 352:180-184(1994).

571. (RrnaAD) Ribosomal RNA adenine dimethylases signature

A number of enzymes responsible for the dimethylation of adenosines if ribosomal RNAs (EC 2.1.1.48) have been found [1,2] to be evolutionary related. These enzymes are:

- Bacterial 16S rRNA dimethylase (gene ksgA), which acts in the biogenesis of ribosomes by catalyzing the dimethylation of two adjacent adenosines in the loop of a conserved hairpin near the 3'-end of 16S rRNA. Inactivation of ksgA leads to resistance to the aminoglycoside antibiotic kasugamycin.
- Yeast 18S rRNA dimethylase (gene DIM1), which is functionally similar to ksgA and that dimethylates twin adenosines in the 3'-end of 18S rRNA.
- Bacterial 'erm' methylases. These enzymes confer resistance to macrolide-lincosamide-streptogramin B (MLS) antibiotics - such as erythromycin - by dimethylating the adenine residue at position 2058 of 23S rRNA thus resulting in a reduced affinity between ribosomes and the MLS antibiotics.
- Caenorhabditis elegans hypothetical protein EO2H1.1.

The best conserved regions in these enzymes is located in the N-terminal section and corresponds to a region that is probably involved in S-adenosyl methionine (SAM) binding.

-Consensus pattern: [LIVM]-[LIVMFY]-[DE]-x-G-[STAPV]-G-x-[GA]-x-[LIVMF]-[ST]-x(2)-[LIVM]-x(6)-[LIVMY]-x-[STAGV]-[LIVMFYHC]-E-x-D

[1] van Gemen B., van Knippenberg P.H. (In) Nucleic acid methylation, Clawson G.A., Willis D.B., Weissbach A., Jones P.A., Eds., pp.19-36, Alan R. Liss Inc, New-York, (1990).
[2] Lafontaine D., Delcour J., Glasser A.L., Desgres J., Vandenhaute J. J. Mol. Biol. 241:492-497(1994).

572. (RuBisC0 small) Ribulose bisphosphate carboxylase, small chain. 206 members

573. ATP/GTP-binding site motif A (P-loop) (ras)

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5]. There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such a motif has been noted are listed below: - ATP synthase alpha and beta subunits. - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins. - Dynamins and dynamin-like proteins - Guanylate kinase - Thymidine kinase (- Thymidylate kinase. - Shikimate kinase. - Nitrogenase iron protein family (nifH/frxC) - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran. - ADP-ribosylation factors family - Bacterial dnaA protein - Bacterial recA protein - Bacterial recF protein - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family - Bacterial type II secretion system protein E. Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved foradenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern: [AG]-x(4)-G-K-[ST]

In addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferritin light chain).

[1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).[2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).[3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl.

Acad. Sci. U.S.A. 83:907-911(1986).[4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).[5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).[6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).[7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).[8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).[9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).[10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

GTP-binding nuclear protein ran signature (ras)

Ran (or TC4) is a small abundant nuclear protein that binds and hydrolyzes GTP and which has been implicated in a large number of processes including nucleocytoplasmic transport, RNA synthesis, processing and export and cell cycle checkpoint control [1,2]. Ran is generally included in the RAS 'superfamily' of small GTP-binding proteins [3], but it is only slightly related to the other RAS proteins. It also differs from RAS proteins in that it lacks cysteine residues at its C- terminal and is therefore not subject to prenylation. Instead ran has an acidic C-terminus. It is, however similar to RAS family members in requiring a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. The region of the GTP-binding B motif which, in ran, is perfectly conserved has been selected as a signature pattern.

Consensus pattern: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y- Proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop).

[1] Scheffzek K., Klebe C., Fritz-Wolf K., Kabsch W., Wittinghofer A. Nature 374:378-381(1995).[2] Rush M.G., Drivas G., d'Eustachio P. BioEssays 18:103-112(1996).[3] Valencia A., Chardin P., Wittinghofer A., Sander C. Biochemistry 30:4637-4648(1991).

574. recA signature

The bacterial recA protein [1,2,3,E1] is essential for homologous recombination and recombinational repair of DNA damage. RecA has many activities: it filaments, it binds to single- and double-stranded DNA, itbinds and hydrolyzes ATP, it is also a recombinase and, finally, it interacts with lexA causing its activation and leading to its autocatalytic cleavage. RecA is a protein of about 350 amino-acid residues. Its sequence is very well conserved [3,4,5,E1] among eubacterial species. It is also found in the chloroplast of plants [6]. The best

conserved region, a nonapeptide located in the middle of the sequence which is part of the monomer-monomer interface in a recA filament has been selected as a signature pattern,.

Consensus pattern: A-L-[KR]-[IF]-[FY]-[STA]-[STAD]-[LIVMQ]-R-

[1] Smith K.C., Wang T.-C. V. BioEssays 10:12-16(1989).[2] Lloyd A.T., Sharp P.M. J. Mol. Evol. 37:399-407(1993).[3] Roca A.I., Cox M.M. Prog. Nucleic Acids Res. Mol. Biol. 56:129-223(1997).[4] Karlin S., Weinstock G.M., Brendel V. J. Bacteriol. 177:6881-6893(1995).[5] Eisen J.A. J. Mol. Evol. 41:1105-1123(1995).[6] Cerutti H.D., Osman M., Grandoni P., Jagendorf A.T. Proc. Natl. Acad. Sci. U.S.A. 89:8068-8072(1992).[E1] http://www.tigr.org/~jeisen/RecA/RecA.html

575. Response regulator receiver domain

This domain receives the signal from the sensor partner inComment: bacterial two-component systems. It is usually found N-terminalComment: to a DNA binding effector domain.

[1] Pao GM, Saier MH; J Mol Evol 1995;40:136-154.

576. Ribonucleotide reductase large subunit signature

*Ribonucleotide reductase (EC 1.17.4.1) [1,2] catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. Ribonucleotide reductase is an oligomeric enzyme composed of a large subunit (700 to 1000 residues) and a small subunit (300 to 400 residues). There are regions of similarities in the sequence of the large chain from prokaryotes, eukaryotes and viruses. One of these regions has been selected as a signature pattern.

Consensus pattern: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-[PA]-

[1] Nillson O., Lundqvist T., Hahne S., Sjoberg B.-M. Biochem. Soc. Trans. 16:91-94(1988).[2] Reichard P. Science 260:1773-1777(1993).

577. Ribonuclease T2 family histidine active sites

The fungal ribonucleases T2 from Aspergillus oryzae, M from Aspergillus saitoiand Rh from Rhizopeus niveus are structurally and functionally related 30 Kdglycoproteins [1] that cleave

the 3'-5' internucleotide linkage of RNA via a nucleotide 2',3'-cyclic phosphate intermediates (EC 3.1.27.1).A number of other RNAses have been found to be evolutionary related to these fungal enzymes: - Self-incompatibility [2] in flowering plants is often controlled by a single gene (S-gene) that has several alleles. This gene prevents fertilization by self-pollen or by pollen bearing either of the two S- alleles expressed in the style. The self-incompatibility glycoprotein from several higher plants of the solanaceae family has been shown [2,3] to be a ribonuclease. - Phosphate-starvation induced RNAses LE and LX from tomato [4]. These two enzymes are probably involved in a phosphate-starvation rescue system. - Escherichia coli periplasmic RNAse I (EC 3.1.27.6) (gene rna) [5]. - Aeromonas hydrophila periplasmic RNAse. - Haemophilus influenzae hypothetical protein HI0526.Two histidines residues have been shown [6,7] to be involved in the catalytic mechanism of RNase T2 and Rh. These residues and the region around them arehighly conserved in all the sequence described above. Two signature patterns have been developed, one for each of the two active-site histidines. The second pattern also contains a cysteine which is known to be involved in a disulfide bond.

Consensus pattern: [FYWL]-x-[LIVM]-H-G-L-W-P [H is an active site residue]

Consensus pattern: [LIVMF]-x(2)-[HDGTY]-[EQ]-[FYW]-x-[KR]-H-G-x-C [H is an active site residue] [C is involved in a disulfide bond]

[1] Watanabe H., Naitoh A., Suyama Y., Inokuchi N., Shimada H., Koyama T., Ohgi K., Irie M. J. Biochem. 108:303-310(1990).[2] Haring V., Gray J.E., McClure B.A., Anderson M.A., Clarke A.E. Science 250:937-941(1990).[3] McClure B.A., Haring V., Ebert P.R., Anderson M.A., Simpson R.J., Sakiyama F., Clarke A.E. Nature 342:95957(1989).[4] Loeffler A., Glund K., Irie M. Eur. J. Biochem. 214:627-633(1993).[5] Meador J. III, Kennell D. Gene 95:1-7(1990).[6] Kawata Y., Sakiyama F., Hayashi F., Kyogoku Y. Eur. J. Biochem. 187:255-262(1990).[7] Kurihara H., Mitsui Y., Ohgi K., Irie M., Mizuno H., Nakamura K.T. FEBS Lett. 306:189-192(1992).

578. Ribonucleotide reductase large subunit signature. Ribonucleotide reductase (EC 1.17.4.1) [1,2] catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. Ribonucleotide reductase is an oligomeric enzyme composed of a large subunit (700 to 1000 residues) and a small subunit (300 to 400 residues). There are regions of similarities in the

sequence of the large chain from prokaryotes, eukaryotes and viruses. One of these regions has been developed as a signature pattern.

Consensus pattern: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-[PA]-

[1] Nillson O., Lundqvist T., Hahne S., Sjoberg B.-M. Biochem. Soc. Trans. 16:91-94(1988).[2] Reichard P. Science 260:1773-1777(1993).

579. RNase H

RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers.

580. Eukaryotic putative RNA-binding region RNP-1 signature (rrm)

Many eukaryotic proteins that are known or supposed to bind single-strandedRNA contain one or more copies of a putative RNA-binding domain of about 90amino acids [1,2]. This region has been found in the following proteins: ** Heterogeneous nuclear ribonucleoproteins ** - hnRNP A1 (helix destabilizing protein) (twice). - hnRNP A2/B1 (twice). - hnRNP C (C1/C2) (once). - hnRNP E (UP2) (at least once). - hnRNP G (once). ** Small nuclear ribonucleoproteins ** - U1 snRNP 70 Kd (once). - U1 snRNP A (once). - U2 snRNP B" (once). ** Pre-RNA and mRNA associated proteins ** - Protein synthesis initiation factor 4B (eIF-4B) [3], a protein essential for the binding of mRNA to ribosomes (once). - Nucleolin (4 times). - Yeast single-stranded nucleic acid-binding protein (gene SSB1) (once). - Yeast protein NSR1 (twice). NSR1 is involved in pre-rRNA processing; it specifically binds nuclear localization sequences. - Poly(A) binding protein (PABP) (4 times). ** Others ** - Drosophila sex determination protein Sex-lethal (Sxl) (twice). - Drosophila sex determination protein Transformer-2 (Tra-2) (once). - Drosophila 'elav' protein (3 times), which is probably involved in the RNA metabolism of neurons. - Human paraneoplastic encephalomyelitis antigen HuD (3 times) [4], which is highly similar to elav and which may play a role in neuron-specific RNA processing. - Drosophila 'bicoid' protein (once) [5], a segment-polarity homeobox protein that may also bind to specific mRNAs. - La

antigen (once), a protein which may play a role in the transcription of RNA polymerase III. - The 60 Kd Ro protein (once), a putative RNP complex protein. - A maize protein induced by abscisic acid in response to water stress, which seems to be a RNA-binding protein. - Three tobacco proteins, located in the chloroplast [6], which may be involved in splicing and/or processing of chloroplast RNAs (twice). - X16 [7], a mammalian protein which may be involved in RNA processing in relation with cellular proliferation and/or maturation. - Insulin-induced growth response protein Cl-4 from rat (twice). - Nucleolysins TIA-1 and TIAR (3 times) [8] which possesses nucleolytic activity against cytotoxic lymphocyte target cells. may be involved in apoptosis. - Yeast RNA15 protein, which plays a role in mRNA stability and/or poly-(A) tail length [9].Inside the putative RNA-binding domain there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif), the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation:

xxxxxxx#####xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx#######xxxxxxxxxxxxxxxxxxxxxxxxxx

RNP-2 RNP-1

The RNP-1 motif has been used as a signature pattern for this type of domain.

Consensus pattern: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM] In most cases the residue in position 3 of the pattern is either Tyr or Phe.

[1] Bandziulis R.J., Swanson M.S., Dreyfuss G. Genes Dev. 3:431-437(1989).[2] Dreyfuss G., Swanson M.S., Pinol-Roma S. Trends Biochem. Sci. 13:86-91(1988).[3] Milburn S.C., Hershey J.W.B., Davies M.V., Kelleher K., Kaufman R.J. EMBO J. 9:2783-2790(1990).[4] Szabo A., Dalmau J., Manley G., Rosenfeld M., Wong E., Henson J., Posner J.B., Furneaux H.M. Cell 67:325-333(1991).[5] Rebagliati M. Cell 58:231-232(1989).[6] Li Y., Sugiura M. EMBO J. 9:3059-3066(1990).[7] Ayane M., Preuss U., Koehler G., Nielsen P.J. Nucleic Acids Res. 19:1273-1278(1991).[8] Kawakami A., Tian Q., Duan X., Streuli M., Schlossman S.F., Anderson P. Proc. Natl. Acad. Sci. U.S.A. 89:8681-8685(1992).[9] Minvielle-Sebastia L., Winsor B., Bonneaud N., Lacroute F. Mol. Cell. Biol. 11:3075-3087(1991).

581. Rubredoxin signature

Rubredoxins [1] are small electron-transfer prokaryotic proteins. They contain an iron atom which is ligated by four cysteine residues. Rubredoxins are, in some cases, functionally interchangeable with ferredoxins.

A conserved region that includes two of the cysteine residues that bind the iron atom has been selected as a pattern for these proteins.

Consensus pattern: [LIVM]-x(3)-W-x-C-P-x-C-[AGD] [The two C's bind the iron atom]

In Pseudomonas oleovorans rubredoxin 2 (gene alkG) [2], this pattern is found twice because alkG has two rubredoxin domains.

Rubrerythrin [3], a protein with inorganic pyrophosphatase activity from Desulfovibrio vulgaris possesses a C-terminal rubredoxin-like domain, but this domain is too divergent to be detected by the above pattern.

[1] Berg J.M., Holm R.H.(In) Iron-sulfur proteins, Spiro T.G., Ed., pp1-66, Wiley, New-York, (1982). [2] Kok M., Oldenhuis R., der Linden M.P.G., Meulenberg C.H.C., Kingma J., Witholt B., J. Biol. Chem. 264:5442-5451(1989). [3] van Beeumen J.J., van Driessche G., Liu M.-Y., Le Gall J., J. Biol. Chem. 266:20645-20653(1991).

582. (rvp) Eukaryotic and viral aspartyl proteases active site

Aspartyl proteases, also known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes [1,2,3] known to exist invertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases are: - Vertebrate gastric pepsins A and C (also known as gastricsin). - Vertebrate chymosin (rennin), involved in digestion and used for making cheese. - Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34). - Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma. - Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21). - Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases. -

Yeast barrier pepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone. - Fission yeast sxa1 which is involved in degrading or processing the mating pheromones. Most retroviruses and some plant viruses, such as badnaviruses, encode for anaspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of a polyprotein which is cleaved during the maturation process of the virus. It is generally part of the pol polyprotein and, more rarely, of the gagpolyprotein. Conservation of the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases allows us to develop a single signature pattern for both groups of protease. Consensus pattern: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]- x-[LIVMFSTNC]-x-[LIVMFGTA] [D is the active site residue] – [1] Foltmann B. Essays Biochem. 17:52-84(1981).[2] Davies D.R. Annu. Rev. Biophys. Chem. 19:189-215(1990).[3] Rao J.K.M., Erickson J.W., Wlodawer A. Biochemistry 30:4663-4671(1991).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:105-120(1995).

583. (rvt) Reverse transcriptase (RNA-dependent DNA polymerase)

A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. Number of members: 1233

[1] Medline: 91006031. Origin and evolution of retroelements based upon their reverse transcriptase sequences. Xiong Y, Eickbush TH; EMBO J 1990;9:3353-3362.

584. (S-AdoMet synt) S-adenosylmethionine synthetase signatures

S-adenosylmethionine synthetase (EC 2.5.1.6) is the enzyme that catalyzes theformation of S-adenosylmethionine (AdoMet) from methionine and ATP [1]. AdoMet is an important methyl donor for transmethylation and is also the propylamino donor in polyamine biosynthesis. In bacteria there is a single isoform of AdoMet synthetase (gene metK), there are two in budding yeast (genes SAM1 and SAM2) and in mammals while in plants there is generally a multigene family.The sequence of AdoMet synthetase is highly conserved throughout isozymes and species. Two signature patterns have been selected for this type of enzyme; the

first is a hexapeptide which seems to be involved in ATP-binding; the second is an almost perfectly conserved glycine-rich nonapeptide.

Consensus pattern: G-A-G-D-Q-G-x(3)-G-[FYH]-Sequences known to belong to this class detected by the pattern:

Consensus pattern: G-[GA]-G-[ASC]-F-S-x-K-[DE]

[1] Horikawa S., Sasuga J., Shimizu K., Ozasa H., Tsukada K. J. Biol. Chem. 265:13683-13686(1990).

585. S1 RNA binding domain

The S1 domain occurs in a wide range of RNAComment: associated proteins. It is structurally similarComment: to cold shock protein which binds nucleic acids.Comment: The S1 domain has an OB-fold structure.

[1] Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG; Cell 1997;88:235-242.

586. SAICAR synthetase signatures

Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6) (SAICARsynthetase) catalyzes the seventh step in the de novo purine biosynthetic pathway; the ATP-dependent conversion of 5'-phosphoribosyl-5-aminoimidazole-4-carboxylic acid and aspartic acid to SAICAR [1]. In bacteria (gene purC),fungi (gene ADE1) and plants, SAICAR synthetase is a monofunctional protein;in higher vertebrates it is the N-terminal domain of a bifunctional enzyme that also catalyze phosphoribosylaminoimidazole carboxylase (AIRC) activity. Two conserved regions in the central section of this enzyme have been selected as signature patterns for SAICAR synthetase.

Consensus pattern: [LIVMF](2)-P-[LIVM]-E-x-[LIVM]-[LIVMCA]-R-x(3)-[TA]-G-S-

Consensus pattern: [LIVM]-[LIVMA]-D-x-K-[LIVMFY]-E-F-G

[1] Zalkin H., Dixon J.E. Prog. Nucleic Acid Res. Mol. Biol. 42:259-287(1992).

587. (SCP) Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signatures

A variety of extracellular proteins from eukaryotes have been found to be evolutionary related: - Rodent sperm-coating glycoprotein (SCP), also known as acidic epididymal glycoprotein (AEG) . This protein is thought to be involved in sperm maturation [1]. It is a

protein of about 220 residues and probably contains eight disulfide bonds. - Mammalian testis-specific protein Tpx-1 [2]. Tpx-1 is highly related to SCP's. - Mammalian glioma pathogenesis-related protein (GliPR). - Lizard helothermine, a toxin that blocks ryanodine receptors. - Venom allergen 5 (Ag5) from vespid wasps and venom allergen 3 (Ag3) from fire ants. These proteins are potent allergens and are the main cause of allergic reactions to stings from insects of the hymenoptera family [3]. Ag5/3 are proteins of about 200 residues and contain four disulfide bonds. - Plant pathogenesis proteins of the PR-1 family [4]. These proteins are synthesized during pathogen infection or other stress-related responses. They are proteins of about 130 to 140 residues and probably contain three disulfide bonds. - Proteins Sc7 and Sc14 from the basidomycete fungus Schizophyllum commune. These extracellular proteins are loosely associated with fruit body hyphal walls [5]. Sc7/14 are proteins of about 180 residues and probably contain two disulfide bonds. - Ancylostoma secreted protein from dog hookworm. - Yeast hypothetical proteins YJL078c, YJL079c and YKR013w.The exact function of these proteins is not yet known. Two conserved regions located in their C-terminal half have been selected as signature patterns. The second signature contains a cysteine which is known to be involved in a disulfide bond in Ag5.

Consensus pattern: [GDER]-H-[FYWH]-T-Q-[LIVM](2)-W-x(2)-[STN]

Consensus pattern: [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N-[LIVMFYWDN] [C is involved in a disulfide bond]

[1] Mizuki N., Kasahara M. Mol. Cell. Endocrinol. 89:25-32(1992).[2] Kasahara M., Gutknecht J., Brew K., Spurr N., Goodfellow P.N. Genomics 5:527-534(1989).[3] Lu G., Villalba M., Coscia M.R., Hoffman D.R., King T.P. J. Immunol. 150:2823-2830(1993).[4] Dixon D.C., Cutt J.R., Klessig D.F. EMBO J. 10:1317-1324(1991).[5] Schuren F.H.J., Asgeirsdottir S.A., Kothe E.M., Scheer J.M.J., Wessels J.G.H. J. Gen. Microbiol. 139:2083-2090(1993).

588. SET domain

SET domains appear to be protein-protein interactionComment: domains. It has been demonstrated that SET domainsComment: mediate interactions with a family of proteins thatComment: display similarity with dual-specificity phosphatasesComment: (dsPTPases) [2].

[1] Tripoulas N, LaJeunesse D, Gildea J, Shearn A; Genetics 1996;143:913-928. [2] Cui X, De Vivo I, Slany R, Miyamoto A, Firestein R, Cleary, ML; Nat Genet 1998;18:331-337.

589. Src homology 3 (SH3) domain profile

The Src homology 3 (SH3) domain is a small protein domain of about 60 amino-acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g. Src, Abl, Lck) [1].Since then, it has been found in a great variety of other intracellular or membrane-associated proteins [2,3,4,5].The SH3 domain has a characteristic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices [6].The function of the SH3 domain is not well understood. The current opinion is that they mediate assembly of specific protein complexes via binding to proline-rich peptides [7].In general SH3 domains are found as single copies in a given protein, but there is a significant number of protein with two SH3 domains and a few with 3 or 4 copies. So far, SH3 domains have been identified in the following proteins: - Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases. - Mammalian phosphatidylinositol-specific phospholipase C-gamma-1 and -2. - Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit. - Mammalian Ras GTPase-activating protein (GAP). - Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK. All of which have two SH3 domains. - Mammalian Vav oncoprotein, a guanine nucleotide exchange factor of the CDC24 family. - Some guanine-nucleotide releasing factors of the CDC25 family: yeast CDC25, yeast SCD25, fission yeast ste6. - MAGUK proteins. These proteins consist of at least three types of domains: one or more copies of the DHR domain, a SH3 domain and a C-terminal guanylate kinase domain. Members of this family are: Drosophila lethal(1)discs large-1 tumor suppressor protein (gene Dlg1), mammalian tight junction protein ZO-1, vertebrate erythrocyte membrane protein p55, Caenorhabditis elegans protein lin-2, rat protein CASK and mammalian synaptic proteins SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1 and SAP102. - Miscellanous proteins interacting with vertebrate receptor protein tyrosine kinases: mammalian cytoplasmic protein Nck (3 copies), oncoprotein Crk (2 copies). - Chicken Src substrate p80/85 protein (cortactin) and the similar human hemopoietic lineage cell specific protein Hs1. - Mammalian dihydrouridine-sensitive L-type calcium channel beta (regulatory) subunit including the related human myasthenic syndrome antigen B (MSYB). - Mammalian neutrophil cytosolic activators of NADPH oxidase: p47 (NCF-1), p67 (NCF-2), and a

potential homolog from Caenorhabditis elegans (B0303.7). NCF-1 and -2 have two copies of the SH3 domain, while B0303.7 has four. - Some myosin heavy chains from amoebae, slime molds and yeast (gene MYO3). - Vertebrate and Drosophila spectrin and fodrin alpha-chain. - Human amphiphysin. - Yeast actin-binding protein ABP1. - Yeast actin-binding protein SLA1 (3 copies). - Yeast protein BEM1 and the fission yeast homolog scd2 (or ral3) (2 copies). - Yeast BEM1-binding proteins BOI2 (BEB1) and BOB1 (BOI1). - Yeast fusion protein FUS1. - Yeast protein RSV167. - Yeast protein SSU81. - Yeast hypothetical proteins YAR014c (1 copy), YFR024c (1 copy), YHL002w (1 copy), YHR016c (1 copy), YJL020C (1 copy), YHR114w (2 copies) and the fission yeast homolog SpAC12C2.05c. - Caenorhabditis elegans hypothetical proteins F42H10.3. The profile developed to detect SH3 domains is based on a structural alignment consisting of 5 gap-free blocks and 4 linker regions totaling 62 match positions.

[1] Mayer B.J., Hamaguchi M., Hanafusa H. Nature 332:272-275(1988).[2] Musacchio A., Gibson T., Lehto V.P., Saraste M. FEBS Lett. 307:55-61(1992).[3] Pawson T., Schlessinger J. Curr. Biol. 3:434-442(1993).[4] Mayer B.J., Baltimore D. Trends Cell Biol. 3:8-13(1993).[5] Pawson T. Nature 373:573-580(1995).[6] Kuriyan J., Cowburn D. Curr. Opin. Struct. Biol. 3:828-837(1993).[7] Morton C.J., Campbell I.D. Curr. Biol. 4:615-617(1994).

590. Serine hydroxymethyltransferase pyridoxal-phosphate attachment site (SHMT)

Serine hydroxymethyltransferase (EC 2.1.2.1) (SHMT) [1] catalyzes the transfer of the hydroxymethyl group of serine to tetrahydrofolate to form 5,10-methylenetetrahydrofolate and glycine. In vertebrates, it exists in acytoplasmic and a mitochondrial form whereas only one form is found in prokaryotes. Serine hydroxymethyltransferase is a pyridoxal-phosphate containing enzyme. The pyridoxal-P group is attached to a lysine residue around which the sequence is highly conserved in all forms of the enzyme.

Consensus pattern: [DEH]-[LIVMFY]-x-[STMV]-[GST]-[ST](2)-H-K-[ST]-[LF]-x-G-[PAC]-[RQ]-[GSA]-[GA] [K is the pyridoxal-P attachment site]

[1] Usha R., Savithri H.S., Rao N.A. Biochim. Biophys. Acta 1204:75-83(1994).

591. SIS domain

SIS (Sugar ISomerase) domains are found in many phosphosugar isomerases and phosphosugar binding proteins.

[1] Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B, Polikarpov I; Structure 1998;6:1047-1055.

592. (SKI) Shikimate kinase signature

Shikimate kinase (EC 2.7.1.71) catalyzes the fifth step in the biosynthesis from chorismate of the aromatic amino acids (the shikimate pathway) inbacteria (gene aroK or aroL), plants and in fungi (where it is part of a multifunctional enzyme which catalyzes five consecutive steps in this pathway).Shikimate kinase is a small protein of about 200 residues. A conserved region that contains a run of three glycines has been selected as a signature pattern. Consensus pattern: [KR]-x(2)-E-x(3)-[LIVMF]-x(8,12)-[LIVMF](2)-[SA]-x-G(3)- x-[LIVMF]. Proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop).

593. SNAP-25 family

SNAP-25 (synaptosome-associated protein 25 kDa) proteins are components of SNARE complexes. Members of this family contain a cluster of cysteine residues that can be palmitoylated for membrane attachment [2].

[1]Brennwald P, Kearns B, Champion K, Keranen S, Bankaitis V, Novick P; Cell 1994;79:245-258. [2] Risinger C, Blomqvist AG, Lundell I, Lambertsson A, Nassel D, Pieribone VA, Brodin L, Larhammar D; J Biol Chem 1993;268:24408-24414.

594. SNF2 and others N-terminal domain

This domain is found in proteins involved in a variety of processes including transcription regulation (e.g., SNF2, STH1, brahma, MOT1) , DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g., lodestar, ETL1).

595. Staphylococcal nuclease homologues (Snase)

Present in all three domains of cellular life. Four copies in the transcriptional coactivator p100. These, however, appear to lack the active site residues of Staphylococcal nuclease. Positions 14 (Asp-21), 34 (Arg-35), 39 (Asp-40), 42 (Glu-43) andComment: 110 (Arg-87) [SNase numbering in parentheses] are thought to be involved in substrate-binding and catalysis.

[1] Ponting CP; Protein Sci 1997;6:459-463. [2] Callebaut I, Mornon JP; Biochem J 1997;321:125-132.

596. SPRY domainA

SPRY Domain is named from SPla and the RYanodine Receptor. Domain of unknown function. Distant homologues are domains in Comment: butyrophilin/marenostrin/pyrin homologues.

[1] Ponting C, Schultz J, Bork P; Trends Biochem Sci 1997;22:193-194.

597. (SQS PSY) Squalene and phytoene synthases signatures

Two different polyisoprene synthases have been shown [1,2,3] to share a number of regions of sequence similarities: - Squalene synthase (EC 2.5.1.21) (farnesyl-diphosphate farnesyltransferase) (SQS), which catalyzes the conversion of two molecules of farnesyl diphosphate (FPP) into squalene. It is the first committed step in the cholesterol biosynthetic pathway. The reaction carried out by SQS is catalyzed in two separate steps: the first is a head-to-head condensation of the two molecules of FPP to form presqualene diphosphate; this intermediate is then rearranged in a NADP-dependent reduction, to form squalene. SQS is found in eukaryotes. In yeast it is encoded by the ERG9 gene, in mammals by the FDFT1 gene. SQS seems to be membrane-bound. - Phytoene synthase (EC 2.5.1.-) (PSY), which catalyzes the conversion of two molecules of geranylgeranyl diphosphate (GGPP) into phytoene. It is the second step in the biosynthesis of carotenoids from isopentenyl diphosphate. The reaction carried out by PSY is catalyzed in two separate steps: the first is a head-to-head condensation of the two molecules of GGPP to form prephytoene diphosphate; this intermediate is then rearranged to form phytoene. PSY is found in all organisms that synthesize carotenoids: plants and photosynthetic bacteria as well as some non-photosynthetic bacteria and fungi. In bacteria PSY is encoded by the gene crtB. In plants PSY

is localized in the chloroplast. As it can be seen from the description above, both SQS and PSY share a number of functional similarities which are also reflected at the level of their primary structure. In particular three well conserved regions are shared bySQS and PSY; they could be involved in substrate binding and/or the catalytic mechanism. Signature patterns have been developed for the second and third conserved regions; they are localized in the central part of these enzymes.

Consensus pattern: Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]- x(2)-[LIV]

Consensus pattern: [LIVM]-G-x(3)-Q-x(2,3)-N-[IF]-x-R-D-[LIVMFY]-x(2)-[DE]- x(4,7)-R-x-[FY]-x-P-

[1] Summers C., Karst F., Charles A.D. Gene 136:185-192(1993).[2] Robinson G.W., Tsay Y.H., Kienzle B.K., Smith-Monroy C.A., Bishop R.W. Mol. Cell. Biol. 13:2706-2727(1993).[3] Roemer S., Hugueney P., Bouvier F., Camara B., Kuntz M. Biochem. Biophys. Res. Commun. 196:1414-1421(1993).

598. SRP54-type proteins GTP-binding domain signature

The signal recognition particle (SRP) is an oligomeric complex that mediates targeting and insertion of the signal sequence of exported proteins into the membrane of the endoplasmic reticulum. SRP consists of a 7S RNA and six protein subunits. One of these subunits, the 54 Kd protein (SRP54), is a GTP-binding protein that interacts with the signal sequence when it emerges from the ribosome. The N-terminal 300 residues of SRP54 include the GTP-binding site (G-domain) and are evolutionary related to similar domains in other proteins which are listed below [1]. - Escherichia coli and Bacillus subtilis ffh protein (P48), a protein which seems to be the prokaryotic counterpart of SRP54. Ffh is associated with a 4.5S RNA in the prokaryotic SRP complex. - Signal recognition particle receptor alpha subunit (docking protein), an integral membrane GTP-binding protein which ensures, in conjunction with SRP, the correct targeting of nascent secretory proteins to the endoplasmic reticulum membrane. The G-domain is located at the C-terminal extremity of the protein. - Bacterial ftsY protein, a protein which is believed to play a similar role to that of the docking protein in eukaryotes. The G-domain is located at the C-terminal extremity of the protein. - The pilA protein from Neisseria gonorrhoeae which seems to be the homolog of ftsY. - A protein from the archaebacteria Sulfolobus solfataricus. This protein is also believed to be a docking protein. The G-domain is also at the C- terminus. - Bacterial flagellar biosynthesis protein flhF. The

best conserved regions in those domains are the sequence motifs that are part of the GTP-binding site, but as those regions are not specific to these proteins, they were not used as a signature pattern. Instead, a conserved region located at the C-terminal end of the domain was selected.

Consensus pattern: P-[LIVM]-x-[FYL]-[LIVMAT]-[GS]-x-[GS]-[EQ]-x(4)-[LIVMF]

[1] Althoff S., Selinger D., Wise J.A. Nucleic Acids Res. 22:1933-1947(1994).

599. (STphosphatase) Serine/threonine specific protein phosphatases signature

Serine/threonine specific protein phosphatases (EC 3.1.3.16) (PP) [1,2,3] are enzymes that catalyze the removal of a phosphate group attached to a serine or evolutionary related. - Protein phosphatase-1 (PP1) is an enzyme of broad specificity. It is inhibited by two thermostable proteins, inhibitor-1 and -2. In mammals, there are two closely related isoforms of PP-1: PP-1alpha and PP-1beta, produced by alternative splicing of the same gene. In Emericella nidulans, PP-1 (gene bimG) plays an important role in mitosis control by reversing the action of the nimA kinase. In yeast, PP-1 (gene SIT4) is involved in dephosphorylating the large subunit of RNA polymerase II. - Protein phosphatase-2A (PP2A) is also an enzyme of broad specificity. PP2A is a trimeric enzyme that consist of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit and a third variable subunit. In mammals, there are two closely related isoforms of the catalytic subunit of PP2A: PP2A-alpha and PP2A-beta, encoded by separate genes. - Protein phosphatase-2B (PP2B or calcineurin), a calcium-dependent enzyme whose activity is stimulated by calmodulin. It is composed of two subunits: the catalytic A-subunit and the calcium-binding B-subunit. The specificity of PP2B is restricted.In addition to the above-mentioned enzymes, some additional serine/threoninespecific protein phosphatases have been characterized and are listed below. - Mammalian phosphatase-X (PP-X), and Drosophila phosphatase-V (PP-V) which are closely related but yet distinct from PP2A. - Yeast phosphatase PPH3, which is similar to PP2A, but with different enzymatic properties. - Drosophila phosphatase-Y (PP-Y), and yeast phosphatases Z1 and Z2 (genes PPZ1 and PPZ2) which are closely related but yet distinct from PP1. - Drosophila retinal degeneration protein C (gene rdgC), a calcium-binding phosphatase required to prevent light-induced retinal degeneration. - Phages Lambda and Phi-80 ORF-221 which have been shown to have phosphatase activity and are related to mammalian PP's. The best conserved regions in these proteins is a perfectly conserved pentapeptide that can be used as a signature pattern.

Consensus pattern: [LIVM]-R-G-N-H-E-

[1] Cohen P. Annu. Rev. Biochem. 58:453-508(1989).[2] Cohen P., Cohen P.T.W. J. Biol. Chem. 264:21435-21438(1989).[3] Cohen P.T.W., Brewis N.D., Hughes V., Mann D.J. FEBS Lett. 268:355-359(1990).

600. Translation initiation factor SUI1 signature

In budding yeast (Saccharomyces cerevisiae), SUI1 is a translation initiation factor that functions in concert with eIF-2 and the initiator tRNA-Met in directing the ribosome to the proper start site of translation [1]. SUI1 is a protein of 108 residues. Close homologs of SUI1 have been found [2] in mammals, insects and plants. SUI1 is also evolutionary related to hypothetical proteins from Escherichia coli (yciH), Haemophilus influenzae (HI1225) and Methanococcus vannielii. A conserved region in the C-terminal section has been selected as a signature pattern.

Consensus pattern: [LIVM]-[EQ]-[LIVM]-Q-G-[DEN]-[KHQ]-[KRV]

[1] Yoon H., Donahue T.F. Mol. Cell. Biol. 12:248-260(1992).[2] Fields C.A., Adams M.D. Biochem. Biophys. Res. Commun. 198:288-291(1994).

601. (S T dehydratase) Serine/threonine dehydratases pyridoxal-phosphate attachment site

Serine and threonine dehydratases [1,2] are functionally and structurally related pyridoxal-phosphate dependent enzymes: - L-serine dehydratase (EC 4.2.1.13) and D-serine dehydratase (EC 4.2.1.14) catalyze the dehydratation of L-serine (respectively D-serine) into ammonia and pyruvate. - Threonine dehydratase (EC 4.2.1.16) (TDH) catalyzes the dehydratation of threonine into alpha-ketobutarate and ammonia. In Escherichia coli and other microorganisms, two classes of TDH are known to exist. One is involved in the biosynthesis of isoleucine, the other in hydroxamino acid catabolism.Threonine synthase (EC 4.2.99.2) is also a pyridoxal-phosphate enzyme, it catalyzes the transformation of homoserine-phosphate into threonine. It has been shown [3] that threonine synthase is distantly related to the serine/threonine dehydratases. In all these enzymes, the pyridoxal-phosphate group is attached to a lysine residue. The sequence around this residue is sufficiently conserved to allow the derivation of a pattern specific to serine/threonine dehydratases and threonine synthases.

Consensus pattern: [DESH]-x(4,5)-[STVG]-x-[AS]-[FYI]-K-[DLIFSA]-[RVMF]-[GA]-[LIVMGA] [The K is the pyridoxal-P attachment site]

[1] Ogawa H., Gomi T., Konishi K., Date T., Naakashima H., Nose K., Matsuda Y., Peraino C., Pitot H.C., Fujioka M. J. Biol. Chem. 264:15818-15823(1989).[2] Datta P., Goss T.J., Omnaas J.R., Patil R.V. Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).[3] Parsot C. EMBO J. 5:3013-3019(1986).[4] Grabowski R., Hofmeister A.E.M., Buckel W. Trends Biochem. Sci. 18:297-300(1993).

Cysteine synthase/cystathionine beta-synthase P-phosphate attachment site

Cysteine synthase (CSase) is the pyridoxal-phosphate dependent enzyme responsible [1] for the formation of cysteine from O-acetyl-serine and hydrogen sulfide with the concomitant release of acetic acid. In bacteria suchas Escherichia coli, two forms of the enzyme are known (genes cysK and cysM).In plants there are also two forms, one located in the cytoplasm and the otherin chloroplasts.Cystathionine beta-synthase [2] catalyzes the first irreversiblestep in homocysteine transulfuration; the conjugation of homocysteine andserine forming cystathionine. Like Csase it is a pyridoxal-phosphate dependent enzyme. The two types of enzymes are evolutionary related. The pyridoxal-phosphategroup of CSases has been shown to be attached to a lysine residue which is located in the N-terminal section of these enzymes; the sequence around this residue is highly conserved and can be used as a signature pattern to detect this class of enzymes.

Consensus pattern: K-x-E-x(3)-[PA]-[STAGC]-x-S-[IVAP]-K-x-R-x-[STAG]-x(2)- [LIVM] [The 2nd K is the pyridoxal-P attachment site

[1] Saito K., Kurosawa M., Murakoshi I. FEBS Lett. 328:111-114(1993).[2] Swaroop M., Bradley K., Ohura T., Tahara T., Roper M.D., Rosenberg L.E., Kraus J.P. J. Biol. Chem. 267:11455-11461(1992).

602. S locus glycop

S-locus glycoprotein family. In Brassicaceae, self-incompatible plants have a self/non-self Comment: recognition system. This is sporophytically controlled by Comment: multiple alleles at a single locus (S). S-locus glycoproteins,Comment: as well as S-receptor kinases, are in linkage with the S-alleles [1].Number of members: 128

[1] Evolutionary aspects of the S-related genes of the Brassica self-incompatibility system: synonymous and nonsynonymous base substitutions. Hinata K, Watanabe M, Yamakawa S,

Satta Y, Isogai A; Genetics 1995;140:1099-1104. [2] Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gene (SLR1) in Raphanus sativus L. and self-incompatible ornamental plants in the Brassicaceae. Sakamoto K, Kusaba M, Nishio T; Mol Gen Genet 1998;258:397-403.

603. (sdh cyt) Succinate dehydrogenase cytochrome b subunit signatures

Succinate dehydrogenase (SDH) is a membrane-bound complex of two main components: a membrane-extrinsic component composed of an FAD-binding flavoprotein and an iron-sulfur protein, and a hydrophobic component composed of a cytochrome B and a membrane anchor protein. The cytochrome b component is a mono heme transmembrane protein [1,2,3] belonging to a family that groups: - Cytochrome b-556 from bacterial SDH (gene sdhC). - Cytochrome b560 from the mammalian mitochondrial SDH complex. - Cytochrome b560 subunit encoded in the mitochondrial genome of some algae and in the plant Marchantia polymorpha. - Cytochrome b from yeast mitochondrial SDH complex (gene SDH3 or CYB3). - Protein cyt-1 from Caenorhabditis.These cytochromes are proteins of about 130 residues that comprise threetransmembrane regions. There are two conserved histidines which may beinvolved in binding the heme group. Two signature patterns have been developed that include these histidine residues.

Consensus pattern: R-P-[LIVMT]-x(3)-[LIVM]-x(6)-[LIVMWPK]-x(4)-S-x(2)-H-R-x- [ST] [H could be a heme ligand]

Consensus pattern: H-x(3)-[GA]-[LIVMT]-R-[HF]-[LIVMF]-x-[FYWM]-D-x-[GVA] [H could be a heme ligand]

[1] Yu L., Wei Y.-Y., Usui S., Yu C.-A. J. Biol. Chem. 267:24508-24515(1992).[2] Abraham P.R., Mulder A., Van't Riet J., Raue H.A. Mol. Gen. Genet. 242:708-716(1994).[3] Leblanc C., Boyen C., Richard O., Bonnard G., Grienenberger J.M., Kloareg B. J. Mol. Biol. 250:484-495(1995).

604. Sec1 family

[1] The Sec1 family: a novel family of proteins involved in synaptic transmission and general secretion. Halachmi N, Lev Z; J Neurochem 1996;66:889-897.

Number of members: 40

605. Protein secE/sec61-gamma signature

In bacteria, the secE protein plays a role in protein export; it is one of the components - with secY and secA - of the preprotein translocase. In eukaryotes, the evolutionary related protein sec61-gamma playsa role in protein translocation through the endoplasmic reticulum; it is part of a trimeric complex that also consist of sec61-alpha and beta [1]. Both secE and sec61-gamma are small proteins of about 60 to 90 amino acids that contain a single transmembrane region at their C-terminal extremity (Escherichia colisecE is an exception, in that it possess an extra N-terminal segment of 60residues that contains two additional transmembrane domains).The sequence of secE/sec61-gamma is not extremely well conserved, however it is possible to derive a signature pattern centered on a conserved proline located 10 residues before the beginning of the transmembrane domain.

Consensus pattern: [LIVMFY]-x(2)-[DENQGA]-x(4)-[LIVMFTA]-x-[KRV]-x(2)-[KW]-P-x(3)-[SEQ]-x(7)-[LIVT]-[LIVGA]-[LIVFGAST]

[1] Hartmann E., Sommer T., Prehn S., Goerlich D., Jentsch S., Rapoport T.A. Nature 367:654-657(1994).

606. 11-S plant seed storage proteins signature

Plant seed storage proteins, whose principal function appears to be the major nitrogen source for the developing plant, can be classified, on the basis of their structure, into different families. 11-S are non-glycosylated proteins which form hexameric structures [1,2]. Each of the subunits in the hexamer is itself composed of an acidic and a basic chain derived from a single precursor and linked by a disulfide bond. This structure is shown in the following representation. +-------------------------+ | | xxxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxNGxCxxxxxxxxxxxxxxxxxxxxxxx ********* <------Acidic-subunit-------------><-----Basic-subunit------> <-----------------About-480-to-500-residues----------------->'C': conserved cysteine involved in a disulfide bond.'*': position of the pattern. Proteins that belong to the 11-S family are: pea and broad bean legumins, rape cruciferin, rice glutelins, cotton beta-globulins, soybean glycinins, pumpkin 11-S globulin, oat globulin, sunflower helianthinin G3, etc. The region that includes the conserved cleavage site between the acidic and basic subunits (Asn-Gly) and a proximal cysteine residue which is involved in the interchain disulfide bond have been used as a signature pattern for this family of proteins.

Consensus pattern: N-G-x-[DE](2)-x-[LIVMF]-C-[ST]-x(11,12)-[PAG]-D [C is involved in a disulfide bond

[1] Hayashi M., Mori H., Nishimura M., Akazawa T., Hara-Nishimura I. Eur. J. Biochem. 172:627-632(1988).[2] Shotwell M.A., Afonso C., Davies E., Chesnut R.S., Larkins B.A. Plant Physiol. 87:698-704(1988).

607. 7S seed storage protein

7S globulin is one of the main storage proteins of most angiosperms and gymnosperms. The 7S storage proteins are homotrimers.

Number of members: 67

[1] The three-dimensional structure of canavalin from jack bean (Canavalia ensiformis). Ko TP, Ng JD, McPherson A; Plant Physiol 1993;101:729-744.

608. Aspartate-semialdehyde dehydrogenase signature

Aspartate-semialdehyde dehydrogenase (ASD) catalyzes the second step in the common biosynthetic pathway leading from Asp to diaminopimelate and Lys, to Met, and to Thr; the NADP-dependent reductive dephosphorylation of L-aspartyl phosphate to L-aspartate-semialdehyde. In bacteria and fungi, ASDis a protein of about 40 Kd (340 to 370 residues) whose sequence is not extremely well conserved [1]. A conserved cysteine residue has been implicated as important for the catalytic activity [2].The region of conservation around the active site residue is too small to be used as signature pattern. Another more conserved region, located in the last third of the sequence, and which contains both a conserved cysteine as well as an histidine has been used instead.

Consensus pattern: [LIVM]-[SADN]-x(2)-C-x-R-[LIVM]-x(4)-[GSC]-H-[STA

[1] Baril C., Richaud C., Fourni E., Baranton G., Saint Girons I. J. Gen. Microbiol. 138:47-53(1992).[2] Karsten W.E., Viola R.E. Biochim. Biophys. Acta 1121:234-238(1992).

N-acetyl-gamma-glutamyl-phosphate reductase active site

N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) [1,2] is the enzyme that catalyzes the third step in the biosynthesis of arginine from glutamate, the NADP-dependent reduction of N-acetyl-5-glutamyl phosphate into N-acetylglutamate 5-semialdehyde.In bacteria it is a monofunctional protein of 35 to 38 Kd (gene argC) while in

fungi it is part of a bifunctional mitochondrial enzyme (gene ARG5,6, arg11 orarg-6) which contains a N-terminal acetylglutamate kinase (EC 2.7.2.8) domain and a C-terminal AGPR domain. In the Escherichia coli enzyme, a cysteine has been shown to be implicated in the catalytic activity, the region around this residue is well conserved and can be used as a signature pattern.

Consensus pattern: [LIVM]-[GSA]-x-P-G-C-[FY]-[AVP]-T-[GA]-x(3)-[GTAC]-[LIVM]- x-P [C is the active site residue]

[1] Ludovice M., Martin J.F., Carrachas P., Liras P. J. Bacteriol. 174:4606-4613(1992).[2] Gessert S.F., Kim J.H., Nargang F.E., Weiss R.L. J. Biol. Chem. 269:8189-8203(1994).

609. Sialyltransferase family,

Number of members: 18

610. SpoU rRNA Methylase family

This family of proteins probably use S-AdoMet. Number of members: 58

[1] SpoU protein of Escherichia coli belongs to a new family of putative rRNA methylases. Koonin EV, Rudd KE; Nucleic Acids Res 1993;21:5519-5519. [2] The spoU gene of escherichia coli , the fourth gene of the spoT operon, is essential for tRNA (Gm18) 2 ' methyltransferase activity. Persson BC, Jager G, Gustafsson C; Nucleic Acids Res 1997;25:4093-4097.

611. Stathmin family signatures

Stathmin [1] (from the Greek 'stathmos' which means relay), is an ubiquitous intracellular protein, present in a variety of phosphorylated forms and which serves as a relay for diverse second messenger pathways. Its expression and phosphorylation are regulated throughout development and in response to extracellular signals regulating cell proliferation, differentiation and function. Stathmin is a highly conserved protein of 149 amino acid residues. Structurally, it consists of an N-terminal domain of about 45 residues followed by a 78 residue alpha-helical domain consisting of a heptad repeat coiled coil structure and a C-terminal domain of 25 residues. Protein SCG10 is a neuron-specific, membrane-associated protein that accumulates in the growth cones of developing neurons. It is highly similar in its

sequence to stathmin, but differs in that it contains an additional N-terminal hydrophobic segment of 32 residues which is probably responsible for its interaction with membranes. Xenopus protein XB3 is also evolutionary related to stathmin and also contains an additional N-terminal hydrophobic domain [2]. A conserved decapeptide which ends with the first three residues of the coiled coil domain and a second pattern that corresponds to part of the central region of the coiled coil have been selected as signatures for proteins of the stathmin family.

Consensus pattern: P-[KRQ]-[KR](2)-[DE]-x-S-L-[EG]-E-

Consensus pattern: A-E-K-R-E-H-E-[KR]-E-

[1] Sobel A. Trends Biochem. Sci. 16:301-305(1991).[2] Maucuer A., Moreau J., Mechali M., Sobel A. J. Biol. Chem. 268:16420-16429(1993).

612. SUA5/yciO/yrdC family signature. The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast protein SUA5. - Escherichia coli hypothetical protein yciO and HI1198, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yrdC and HI0656, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein ywlC. - Mycobacterium leprae hypothetical protein in rfe-hemK intergenic region. - Methanococcus jannaschii hypothetical protein MJ0062.These are proteins of from 20 to 46 Kd which contain a number of conserved regions in their N-terminal section. They can be picked up in the database by the following pattern.

Consensus pattern: [LIVMTA](3)-[LIVMFYC]-[PG]-T-[DE]-[STA]-x-[FY]-[GA]- [LIVM]-[GS]-

[1] Bairoch A., Rudd K.E., Robison K. Unpublished observations (1995).

613. Sucrose synthase

Sucrose synthases catalyse the synthesis of sucrose from UDP-glucose and fructose. This family includes the bulk of the sucrose synthase protein. However the carboxyl terminal region of the sucrose synthases belongs to the glycosyl transferase family Glycos_transf_1.

614. Sulfotransferase proteins

Number of members: 59

615. Synaptophysin / synaptoporin signature

Synaptophysin and synaptoporin [1] are structurally related proteins, found in the membrane of synaptic vesicles, which may function as ionic or solute channels. These two glycoproteins seem to span the membrane four times. Both their N- and C-termini sequences seem to be cytoplasmically located. As a signature pattern for this family of proteins, a highly conserved region located in the beginning of the first intravesicular loop just after the first transmembrane domain has been selected. This region contains a cysteine residue that may be involved in a disulfide bond.

Consensus pattern: L-S-V-[DE]-C-x-N-K-T [C may be involved in a disulfide bond

[1] Knaus P., Marqueze-Pouey B., Scherer H., Betz H. Neuron 5:453-462(1990).

616. Syndecans signature

Syndecans [1,2] (from the greek syndein; to bind together) are a family of transmembrane heparan sulfate proteoglycans which are implicated in the binding of extracellular matrix components and growth factors. Syndecans bind a variety of molecules via their heparan sulfate chains and can act as receptors or as co-receptors. Structurally, these proteins consist of four separate domains: a) A signal sequence; b) An extracellular domain (ectodomain) of variable length and whose sequence is not evolutionary conserved in the various forms of syndecans. The ectodomain contains the sites of attachment of the heparan sulfate glycosaminoglycan side chains; c) A transmembrane region; d) A highly conserved cytoplasmic domain of about 30 to 35 residues which could interact with cytoskeletal proteins. The proteins known to belong to this family are: - Syndecan 1. - Syndecan 2 or fibroglycan. - Syndecan 3 or neuroglycan or N-syndecan. - Syndecan 4 or amphiglycan or ryudocan. - Drosophila syndecan. - Caenorhabditis elegans probable syndecan (F57C7.3).The signature pattern that has been developed for syndecans starts with the last residue of the transmembrane region and includes the first 10 residues of the cytoplasmic domain. This region, which contains four basic residues, could act as a stop transfer site.

Consensus pattern: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y

[1] Bernfield M., Kokenyesi R., Kato M., Hinkes M.T., Spring J., Gallo R.L., Lose E.J. Annu. Rev. Cell Biol. 8:365-393(1992).[2] David G. FASEB J. 7:1023-1030(1993).

617. Syntaxin / epimorphin family signature

The following proteins have been shown to be evolutionary related [1,2,3]: - Epimorphin (or syntaxin 2), a mammalian mesenchymal protein which plays an essential role in epithelial morphogenesis. - Syntaxin 1A (also known as antigen HPC-1) and syntaxin 1B which are synaptic proteins which may be involved in docking of synaptic vesicles at presynaptic active zones. - Syntaxin 3. - Syntaxin 4, which is potentially involved in docking of synaptic vesicles at presynaptic active zones. - Syntaxin 5, which mediates endoplasmic reticulum to golgi transport. - Syntaxin 6, which is involved in intracellular vesicle trafficking. - Syntaxin 7. - Yeast PEP12 (or VPS6) which is required for the transport of proteases to the vacuole. - Yeast SED5 which is required for the fusion of transport vesicles with the Golgi complex. - Yeast SSO1 and SSO2 which are required for vesicle fusion with the plasma membrane. - Yeast VAM3, which is required for vacuolar assembly. - Arabidopsis thaliana protein KNOLLE which may be involved in cytokinesis. - Caenorhabditis elegans hypothetical proteins F35C8.4, F48F7.2, F55A11.2 and T01B11.3.The above proteins share the following characteristics: a size ranging from30 Kd to 40 Kd; a C-terminal extremity which is highly hydrophobic and isprobably involved in anchoring the protein to the membrane; a central, well conserved region, which seems to be in a coiled-coil conformation. The pattern specific for this family is based on the most conserved region of the coiled coil domain.

Consensus pattern: [RQ]-x(3)-[LIVMA]-x(2)-[LIVM]-[ESH]-x(2)-[LIVMT]-x-[DEVM]-[LIVM]-x(2)-[LIVM]-[FS]-x(2)-[LIVM]-x(3)-[LIVT]-x(2)-Q- [GADEQ]-x(2)-[LIVM]-[DNQT]-x-[LIVMF]-[DESV]-x(2)-[LIVM]

[1] Bennett M.K., Garcia-Arraras J.E., Elferink L.A., Peterson K., Fleming A.M., Hazuka C.D., Scheller R.H. Cell 74:863-873(1993).[2] Spring J., Kato M., Bernfield M. Trends Biochem. Sci. 18:124-125(1993).[3] Pelham H.R.B. Cell 73:425-426(1993).

618. Sm protein

The U1, U2, U4/U6, and U5 small nuclear ribonucleoprotein particles (snRNPs) involved in pre-mRNA splicing contain seven Sm proteins (B/B', D1, D2, D3, E, F and G) in common, which

assemble around the Sm site present in four of the major spliceosomal small nuclear RNAs. These proteins contain a common sequence motif in two segments, Sm1 and Sm2, separated by a short variable linker.

[1] Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Luhrmann R EMBO J 1995;14:2076-2088. [2] Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Luhrmann R, Li J, Nagai K; Cell 1999;96:375-387.

619. Skp1 family

[1] Stebbins CE, Kaelin WG Jr, Pavletich NP; Science 1999;284:455-461.

620. Protein secY signatures

The eubacterial secY protein [1] plays an important role in protein export. It interacts with the signal sequences of secretory proteins as well as with two other components of the protein translocation system: secA and secE. SecY is an integral plasma membrane protein of 419 to 492 amino acid residues that apparently contains ten transmembrane segments. Such a structure probablyconfers to secY a 'translocator' function, providing a channel for periplasmic and outer-membrane precursor proteins.Homologs of secY are found in archaebacteria [2]. SecY is also encoded in the chloroplast genome of some algae [3] where it could be involved in a prokaryotic-like protein export system across the two membranes of the chloroplast endoplasmic reticulum (CER) which is present in chromophyte andcryptophyte algae. Two signature patterns have been developed for secY proteins. The first corresponds to the second transmembrane region, which is the most conserved section of these proteins. The second spans the C-terminal part of the fourth transmembrane region, a short intracellular loop, and the N-terminal part of the fifth transmembrane region.

Consensus pattern: [GST]-[LIVMF](2)-x-[LIVM]-G-[LIVM]-x-P-[LIVMFY](2)-x-[AS]-[GSTQ]-[LIVMFAT](3)-Q-[LIVMFA](2)

Consensus pattern: [LIVMFYW](2)-x-[DE]-x-[LIVMF]-[STN]-x(2)-G-[LIVMF]-[GST]-[NST]-G-x-[GST]-[LIVMF](3)

[1] Ito K. Mol. Microbiol. 6:2423-2428(1992).[2] Auer J., Spicker G., Boeck A. Biochimie 73:683-688(1991).[3] Douglas S.E. FEBS Lett. 298:93-96(1992).

621. (Seed protein) Small hydrophilic plant seed proteins signature. The following small hydrophilic plant seed proteins are structurally related: - Arabidopsis thaliana proteins GEA1 and GEA6. - Cotton late embryogenesis abundant (LEA) protein D-19. - Carrot EMB-1 protein. - Barley LEA proteins B19.1A, B19.1B, B19.3 and B19.4. - Maize late embryogenesis abundant protein Emb564. - Radish late seed maturation protein p8B6. - Rice embryonic abundant protein Emp1. - Sunflower 10 Kd late embryogenesis abundant protein (DS10). - Wheat Em proteins. These proteins contains from 83 to 153 amino acid residues and may play a role[1,2] in equipping the seed for survival, maintaining a minimal level of hydration in the dry organism and preventing the denaturation of cytoplasmic components. They may also play a role during imbibition by controlling water uptake. As a signature pattern, the best conserved region in the sequence of these proteins has been developed, it is a glycine-rich nonapeptide located in the N-terminal section.-

Consensus pattern: G-[EQ]-T-V-V-P-G-G-T-

[1] Dure L. III, Crouch M., Harada J., Ho T.-H. D., Mundy J., Quatrano R., Thomas T., Sung Z.R. Plant Mol. Biol. 12:475-486(1989).[2] Gaubier P., Raynal M., Hull G., Huestis G.M., Grellet F., Arenas C., Pages M., Delseny M. Mol. Gen. Genet. 238:409-418(1993).

622. Serine carboxypeptidases, active sites

All known carboxypeptidases are either metallo carboxypeptidases or serinecarboxypeptidases. The catalytic activity of the serine carboxypeptidases, like that of the trypsin family serine proteases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine [1]. Proteins known to be serine carboxypeptidases are: - Barley and wheat serine carboxypeptidases I, II, and III [2]. - Yeast carboxypeptidase Y (YSCY) (gene PRC1), a vacuolar protease involved in degrading small peptides. - Yeast KEX1 protease, involved in killer toxin and alpha-factor precursor processing. - Fission yeast sxa2, a probable carboxypeptidase involved in degrading or processing mating pheromones [3]. - Penicillium

janthinellum carboxypeptidase S1 [4]. - Aspergullus niger carboxypeptidase pepF. - Aspergullus satoi carboxypeptidase cpdS. - Vertebrate protective protein / cathepsin A [5], a lysosomal protein which is not only a carboxypeptidase but also essential for the activity of both beta-galactosidase and neuraminidase. - Mosquito vitellogenic carboxypeptidase (VCP) [6]. - Naegleria fowleri virulence-related protein Nf314 [7]. - Yeast hypothetical protein YBR139w. - Caenorhabditis elegans hypothetical proteins C08H9.1, F13D12.6, F32A5.3, F41C3.5 and K10B2.2.This family also includes: - Sorghum (s)-hydroxymandelonitrile lyase (hydroxynitrile lyase) (HNL) [8], an enzyme involved in plant cyanogenesis. The sequences surrounding the active site serine and histidine residues are highly conserved in all these serine carboxypeptidases.

Consensus pattern: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS] [S is the active site residue]

Consensus pattern: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x-[IVAQ]-P-x(3)-[PSA] [H is the active site residue]

[1] Liao D.I., Remington S.J. J. Biol. Chem. 265:6528-6531(1990).[2] Sorensen S.B., Svendsen I., Breddam K. Carlsberg Res. Commun. 54:193-202(1989).[3] Imai Y., Yamamoto M. Mol. Cell. Biol. 12:1827-1834(1992).[4] Svendsen I., Hofmann T., Endrizzi J., Remington J., Breddam K. FEBS Lett. 333:39-43(1993).[5] Galjart N.J., Morreau H., Willemsen R., Gillemans N., Bonten E.J., d'Azzo A. J. Biol. Chem. 266:14754-14762(1991).[6] Cho W.L., Deitsch K.W., Raikhel A.S. Proc. Natl. Acad. Sci. U.S.A. 88:10821-10824(1991).[7] Hu W.N., Kopachik W., Band R.N. Infect. Immun. 60:2418-2424(1992).[8] Wajant H., Mundry K.W., Pfitzenmaier K. Plant Mol. Biol. 26:735-746(1994).[9] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

623. Serpins signature. Serpins (SERine Proteinase INhibitors) [1,2,3,4] are a group of structurally related proteins. They are high molecular weight (400 to 500 amino acids),extracellular, irreversible serine protease inhibitors with a well defined structural-functional characteristic: a reactive region that acts as a 'bait' for an appropriate serine protease. This region is found in the C-terminal part of these proteins. Proteins which are known to belong to the serpin family are listed below (references are only provided for recently determined sequences): - Alpha-1 protease inhibitor (alpha-1-antitrypsin, contrapsin). - Alpha-1-antichymotrypsin, - Antithrombin III. - Alpha-2-antiplasmin. - Heparin cofactor II. - Complement C1 inhibitor. - Plasminogen activator inhibitors 1 (PAI-1) and 2 (PAI-2). - Glia derived nexin (GDN) (Protease nexin I). - Protein C inhibitor. - Rat

hepatocytes SPI-1, SPI-2 and SPI-3 inhibitors. - Human squamous cell carcinoma antigen (SCCA) which may act in the modulation of the host immune response against tumor cells. - A lepidopteran protease inhibitor. - Leukocyte elastase inhibitor which, in contrast to other serpins, is an intracellular protein. - Neuroserpin [5], a neuronal inhibitor of plasminogen activators and plasmin. - Cowpox virus crmA [6], an inhibitor of the thiol protease interleukin-1B converting enzyme (ICE). CrmA is the only serpin known to inhibit a non-serine proteinase. - Some orthopoxviruses probable protease inhibitors, which may be involved in the regulation of the blood clotting cascade and/or of the complement cascade in the mammalian host. On the basis of strong sequence similarities, a number of proteins with no known inhibitory activity are said to belong to this family: - Birds ovalbumin and the related genes X and Y proteins. - Angiotensinogen; the precursor of the angiotensin active peptide. - Barley protein Z; the major endosperm albumin. - Corticosteroid binding globulin (CBG). - Thyroxine-binding globulin (TBG). - Sheep uterine milk protein (UTMP) and pig uteroferrin-associated protein (UFAP). - Hsp47, an endoplasmic reticulum heat-shock protein that binds strongly to collagen and could act as a chaperone in the collagen biosynthetic pathway [7]. - Maspin, which seems to function as a tumor supressor [5]. - Pigment epithelium-derived factor precursor (PEDF), a protein with a strong neutrophic activity [8]. - Ep45, an estrogen-regulated protein from Xenopus [9]. A signature pattern has been developed for this family of proteins, centered on a well conserved Pro-Phe sequence which is found ten to fifteen residues on the C-terminal side of the reactive bond

Consensus pattern: [LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-[LIVMFAH]-

[1] Carrell R., Travis J. Trends Biochem. Sci. 10:20-24(1985).[2] Carrell R., Pemberton P.A., Boswell D.R. Cold Spring Harbor Symp. Quant. Biol. 52:527-535(1987).[3] Huber R., Carrell R.W. Biochemistry 28:8951-8966(1989).[4] Remold-O'Donneel E. FEBS Lett. 315:105-108(1993).[5] Osterwalder T., Contartese J., Stoeckli E.T., Kuhn T.B., Sonderegger P. EMBO J. 15:2944-2953(1996).[6] Komiyama T., Ray C.A., Pickup D.J., Howard A.D., Thornberry N.A., Peterson E.P., Salvesen G. J. Biol. Chem. 269:19331-19337(1994).[7] Clarke E., Sandwal B.D. Biochim. Biophys. Acta 1129:246-248(1992).[8] Zou Z., Anisowicz A., Neveu M., Rafidi K., Sheng S., Sager R., Hendrix M.J., Seftor E., Thor A. Science 263:526-529(1994).[9] Steele F.R., Chader G.J., Johnson L.V., Tombran-Tink J.

Proc. Natl. Acad. Sci. U.S.A. 90:1526-1530(1993).[10] Holland L.J., Suksang C., Wall A.A., Roberts L.R., Moser D.R., Bhattacharya A. J. Biol. Chem. 267:7053-7059(1992).

624. Sigma-54 interaction domain signatures and profile

Some bacterial regulatory proteins activate the expression of genes from promoters recognized by core RNA polymerase associated with the alternative sigma-54 factor. These have a conserved domain of about 230 residues involved in the ATP-dependent [1,2] interaction with sigma-54. This domain has been found in the proteins listed below: - acoR from Alcaligenes eutrophus, an activator of the acetoin catabolism operon acoXABC. - algB from Pseudomonas aeruginosa, an activator of alginate biosynthetic gene algD. - dctD from Rhizobium, an activator of dctA, the C4-dicarboxylate transport protein. - dhaR from Citrobacter freundii, a regulator of the dha operon for glycerol utilization. - fhlA from Escherichia coli, an activator of the formate dehydrogenase H and hydrogenase III structural genes. - flbD from Caulobacter crescentus, an activator of flagellar genes. - hoxA from Alcaligenes eutrophus, an activator of the hydrogenase operon. - hrpS from Pseudomonas syringae, an activator of hprD as well as other hrp loci involved in plant pathogenicity. - hupR1 from Rhodobacter capsulatus, an activator of the [NiFe] hydrogenase genes hupSL. - hydG from Escherichia coli and Salmonella typhimurium, an activator of the hydrogenase activity. - levR from Bacillus subtilis, which regulates the expression of the levanase operon (levDEFG and sacC). - nifA (as well as anfA and vnfA) from various bacteria, an activator of the nif nitrogen-fixing operon. - ntrC, from various bacteria, an activator of nitrogen assimilatory genes such as that for glutamine synthetase (glnA) or of the nif operon. - pgtA from Salmonella typhimurium, the activator of the inducible phospho- glycerate transport system. - pilR from Pseudomonas aeruginosa, an activator of pilin gene transcription. - rocR from Bacillus subtilis, an activator of genes for arginine utilization - tyrR from Escherichia coli, involved in the transcriptional regulation of aromatic amino-acid biosynthesis and transport. - wtsA, from Erwinia stewartii, an activator of plant pathogenicity gene wtsB. - xylR from Pseudomonas putida, the activator of the tol plasmid xylene catabolism operon xylCAB and of xylS. - Escherichia coli hypothetical protein yfhA. - Escherichia coli hypothetical protein yhgB. About half of these proteins (algB, dcdT, flbD, hoxA, hupR1, hydG, ntrC, pgtA and pilR) belong to signal transduction two-component systems [3] and possess a domain that can be phosphorylated by a sensor-kinase protein in their N- terminal section. Almost all of these proteins possess a helix-turn-helix DNA-binding domain in their

C-terminal section. The domain which interacts with the sigma-54 factor has an ATPase activity. This may be required to promote a conformational change necessary for theinteraction [4]. The domain contains an atypical ATP-binding motif A (P-loop) as well as a form of motif B. The two ATP-binding motifs are located in the N-terminal section of the domain; signature patterns have been developed for both motifs. Other regions of the domain are also conserved. One of them, located in the C-terminal section, has been selected as a third signature pattern.

Consensus pattern: [LIVMFY](3)-x-G-[DEQ]-[STE]-G-[STAV]-G-K-x(2)-[LIVMFY]

Consensus pattern: [GS]-x-[LIVMF]-x(2)-A-[DNEQASH]-[GNEK]-G-[STIM]-[LIVMFY](3)-[DE]-[EK]-[LIVM]

Consensus pattern: [FYW]-P-[GS]-N-[LIVM]-R-[EQ]-L-x-[NHAT]

[1] Morrett E., Segovia L. J. Bacteriol. 175:6067-6074(1993).[2] Austin S., Kundrot C., Dixon R. Nucleic Acids Res. 19:2281-2287(1991).[3] Albright L.M., Huala E., Ausubel F.M. Annu. Rev. Genet. 23:311-336(1989).[4] Austin S., Dixon R. EMBO J. 11:2219-2228(1992).

625. Sigma-70 factors family signatures

Sigma factors [1] are bacterial transcription initiation factors that promote the attachment of the core RNA polymerase to specific initiation sites and arethen released. They alter the specificity of promoter recognition. Most bacteria express a multiplicity of sigma factors. Two of these factors, sigma-70 (gene rpoD), generally known as the major or primary sigma factor, and sigma-54 (gene rpoN or ntrA) direct the transcription of a wide variety of genes. The other sigma factors, known as alternative sigma factors, are required for the transcription of specific subsets of genes. With regard to sequence similarity, sigma factors can be grouped into two classes: the sigma-54 and sigma-70 families. The sigma-70 family includes, in addition to the primary sigma factor, a wide variety of sigma factors, some of which are listed below: - Bacillus sigma factors involved in the control of sporulation-specific genes: sigma-E (sigE or spoIIGB), sigma-F (sigF or spoIIAC), sigma-G (sigG or spoIIIG), sigma-H (sigH or spo0C) and sigma-K (sigK or spoIVCB/spoIIIC). - Escherichia coli and related bacteria sigma-32 (gene rpoH or htpR) involved in the expression of heat shock genes. - Escherichia coli and related bacteria sigma-27 (gene fliA) involved in the expression of the flagellin gene. - Escherichia coli sigma-S (gene rpoS or katF) which seems to be involved in the expression of genes required for protection against external stresses. - Myxococcus xanthus sigma-B

(sigB) which is essential for the late-stage differentiation of that bacteria. Alignments of the sigma-70 family permit the identification of four regions of high conservation [2,3]. Each of these four regions can in turn be subdivided into a number of sub-regions. Signature patterns based on the two best-conserved sub-regions have been developed. The first pattern corresponds to sub-region 2.2;the exact function of this sub-region is not known although it could be involved in the binding of the sigma factor to the core RNA polymerase. The second pattern corresponds to sub-region 4.2 which seems to harbor a DNA-binding 'helix-turn-helix' motif involved in binding the conserved -35region of promoters recognized by the major sigma factors. The second pattern starts one residue before the N-terminal extremity of the HTH region and ends six residues after its C-terminal extremity.

Consensus pattern: [DE]-[LIVMF](2)-[HEQS]-x-G-x-[LIVMFA]-G-L-[LIVMFYE]-x-[GSAM]-[LIVMAP]

Consensus pattern: [STN]-x(2)-[DEQ]-[LIVM]-[GAS]-x(4)-[LIVMF]-[PSTG]-x(3)-[LIVMA]-x-[NQR]-[LIVMA]-[EQH]-x(3)-[LIVMFW]-x(2)-[LIVM]

[1] Helmann J.D., Chamberlin M.J. Annu. Rev. Biochem. 57:839-872(1988).[2] Gribskov M., Burgess R.R. Nucleic Acids Res. 14:6745-6763(1986).[3] Lonetto M.A., Gribskov M., Gross C.A. J. Bacteriol. 174:3843-3849(1992).[4] Lonetto M.A., Brown K.L., Rudd K.E., Buttner M.J. Proc. Natl. Acad. Sci. U.S.A. 91:7573-7577(1994).

626. Signal carboxyl-terminal domain. 430 members.

627. Signal peptidases I signatures

Signal peptidases (SPases) [1] (also known as leader peptidases) remove the signal peptides from secretory proteins. In prokaryotes three types of Spases are known: type I (gene lepB) which is responsible for the processing of the majority of exported pre-proteins; type II (gene lsp) which only process lipoproteins, and a third type involved in the processing of pili subunits. SPase I is an integral membrane protein that is anchored in the cytoplasmic membrane by one (in B. subtilis) or two (in E. coli) N-terminal transmembrane domains with the main part of the protein protuding in the periplasmic space. Two residues have been shown [2,3] to be essential for the catalytic activity of SPase I: a serine and an lysine.SPase I is evolutionary related to the yeast mitochondrial inner membrane protease subunit 1 and 2 (genes IMP1 and IMP2) which catalyze the removal of signal peptides required for the

targeting of proteins from the mitochondrial matrix, across the inner membrane, into the inter-membrane space [4]. In eukaryotes the removal of signal peptides is effected by an oligomeric enzymatic complex composed of at least five subunits: the signal peptidase complex (SPC). The SPC is located in the endoplasmic reticulum membrane. Two components of mammalian SPC, the 18 Kd (SPC18) and the 21 Kd (SPC21) subunits as well as the yeast SEC11 subunit have been shown [5] to share regions of sequence similarity with prokaryotic SPases I and yeast IMP1/IMP2. Three signature patterns for these proteins have been developed. The first signature contains the putative active site serine, the second signature contains the putative active site lysine which is not conserved in the SPC subunits, and the third signature corresponds to a conserved region of unknown iological significance which is located in the C-terminal section of all these proteins.

Consensus pattern: [GS]-x-S-M-x-[PS]-[AT]-[LF] [S is an active site residue]

Consensus pattern: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY] [K is an active site residue]

Consensus pattern: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG]

[1] Dalbey R.E., von Heijne G. Trends Biochem. Sci. 17:474-478(1992).[2] Sung M., Dalbey R.E. J. Biol. Chem. 267:13154-13159(1992).[3] Black M.T. J. Bacteriol. 175:4957-4961(1993).[4] Nunnari J., Fox T.D., Walter P. Science 262:1997-2004(1993).[5] van Dijl J.M., de Jong A., Vehmaanpera J., Venema G., Bron S. EMBO J. 11:2819-2828(1992).[6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

628. (sodcu) Copper/Zinc superoxide dismutase signatures

Copper/Zinc superoxide dismutase (SODC) [1] is one of the three forms of an enzyme that catalyzes the dismutation of superoxide radicals. SODC binds one atom each of zinc and copper. Various forms of SODC are known: acytoplasmic form in eukaryotes, an additional chloroplast form in plants, an extracellular form in some eukaryotes, and a periplasmic form in prokaryotes. The metal binding sites are conserved in all the known SODC sequences [2]. Two signature patterns have been derived for this family of enzymes: the first one contains two histidine residues that bind the copper atom; the second one islocated in the C-terminal section of SODC and contains a cysteine which is involved in a disulfide bond.

Consensus pattern: [GA]-[IMFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGDE] [The two H's are copper ligands]

Consensus pattern: G-[GN]-[SGA]-G-x-R-x-[SGA]-C-x(2)-[IV] [C is involved in a disulfide bond]

[1] Bannister J.V., Bannister W.H., Rotilio G. CRC Crit. Rev. Biochem. 22:111-154(1987).[2] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:175-184(1992).

629. (sodfe) Manganese and iron superoxide dismutases signature

Manganese superoxide dismutase (SODM) [1] is one of the three forms of an enzyme that catalyzes the dismutation of superoxide radicals. The four ligands of the manganese atom are conserved in all the known SODM sequences. These metal ligands are also conserved in the related iron form ofsuperoxide dismutases [2,3]. A short conserved region which includes two of the four ligands: an aspartate and a histidine has been selected as a signature.

Consensus pattern: D-x-W-E-H-[STA]-[FY](2) [D and H are manganese/iron ligands]

[1] Bannister J.V., Bannister W.H., Rotilio G. CRC Crit. Rev. Biochem. 22:111-154(1987).[2] Parker M.W., Blake C.C.F. FEBS Lett. 229:377-382(1988).[3] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:175-184(1992).

630. Spectrin repeat

Spectrin repeats are found in several proteins involved in cytoskeletal structure. These include spectrin, alpha-actinin and dystrophin.The sequence repeat used in this family is taken from the structural repeat in reference [2]. The spectrin repeat forms a three helix bundle. The second helix is interrupted by proline in some sequences.

Number of members: 898

[1] Actin-binding proteins. 1: Spectrin super family. Hartwig JH; Protein Profile 1995;2:732-732. [2] Crystal structure of the repetitive segments of spectrin. Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D; Science 1993;262:2027-2030.

631. (subtilase) Streptomyces subtilisin-type inhibitors signature

Bacteria of the Streptomyces family produce a family of proteinase inhibitors[1] characterized by their strong activity toward subtilisin. They arecollectively known as SSI's: Streptomyces Subtilisin Inhibitors. Some SSI'salso inhibit trypsin or chymotrypsin. In their

mature secreted form, SSI's areproteins of about 110 residues with two conserved disulfide bonds. +-------+ +--------------+ |||| xxxxxxxxxxxxxxCxxxxxxxCxxxxxxxxxCx#xxxxxxxxxxxxCxxxxxx ************'C': conserved cysteine involved in a disulfide bond.'#': active site residue.'*': position of the pattern.

Consensus pattern: C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L [The two C's are involved in a disulfide bond]

[1] Taguchi S., Kojima S., Terabe M., Miura K.-I., Momose H. Eur. J. Biochem. 220:911-918(1994).

632. Sugar transport proteins signatures

In mammalian cells the uptake of glucose is mediated by a family of closely related transport proteins which are called the glucose transporters [1,2,3].At least seven of these transporters are currently known to exist (in Human they are encoded by the GLUT1 to GLUT7 genes).These integral membrane proteins are predicted to comprise twelve membrane spanning domains. The glucose transporters show sequence similarities [4,5] with a number of other sugar or metabolite transport proteins listed below (references are only provided for recently determined sequences). - Escherichia coli arabinose-proton symport (araE). - Escherichia coli galactose-proton symport (galP). - Escherichia coli and Klebsiella pneumoniae citrate-proton symport (also known as citrate utilization determinant) (gene cit). - Escherichia coli alpha-ketoglutarate permease (gene kgtP). - Escherichia coli proline/betaine transporter (gene proP) [6]. - Escherichia coli xylose-proton symport (xylE). - Zymomonas mobilis glucose facilitated diffusion protein (gene glf). - Yeast high and low affinity glucose transport proteins (genes SNF3, HXT1 to HXT14). - Yeast galactose transporter (gene GAL2). - Yeast maltose permeases (genes MAL3T and MAL6T). - Yeast myo-inositol transporters (genes ITR1 and ITR2). - Yeast carboxylic acid transporter protein homolog JEN1. - Yeast inorganic phosphate transporter (gene PHO84). - Kluyveromyces lactis lactose permease (gene LAC12). - Neurospora crassa quinate transporter (gene Qa-y), and Emericella nidulans quinate permease (gene qutD). - Chlorella hexose carrier (gene HUP1). - Arabidopsis thaliana glucose transporter (gene STP1). - Spinach sucrose transporter. - Leishmania donovani transporters D1 and D2. - Leishmania enriettii probable transport protein (LTP). - Yeast hypothetical proteins YBR241c, YCR98c and YFL040w. - Caenorhabditis elegans hypothetical protein ZK637.1. - Escherichia coli hypothetical proteins

yabE, ydjE and yhjE. - Haemophilus influenzae hypothetical proteins HI0281 and HI0418. - Bacillus subtilis hypothetical proteins yxbC and yxdF. It has been suggested [4] that these transport proteins have evolved from theduplication of an ancestral protein with six transmembrane regions, this hypothesis is based on the conservation of two G-R-[KR] motifs. The first one is located between the second and third transmembrane domains and the second one between transmembrane domains 8 and 9. Two patterns have been developed to detect this family of proteins. The first pattern is based on the G-R-[KR] motif; but because this motif is too short to be specific to this family of proteins, a pattern from a larger region centered on the second copy of this motif was derived. The second pattern is based on a number of conserved residues which are located at the end of the fourth transmembrane segment and in the short loop region between the fourth and fifth segments.

Consensus pattern: [LIVMSTAG]-[LIVMFSAG]-x(2)-[LIVMSA]-[DE]-x-[LIVMFYWA]-G- R-[RK]-x(4,6)-[GSTA]

Consensus pattern: [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)- [RK]

[1] Silverman M. Annu. Rev. Biochem. 60:757-794(1991).[2] Gould G.W., Bell G.I. Trends Biochem. Sci. 15:18-23(1990).[3] Baldwin S.A. Biochim. Biophys. Acta 1154:17-49(1993).[4] Maiden M.C.J., Davis E.O., Baldwin S.A., Moore D.C.M., Henderson P.J.F. Nature 325:641-643(1987).[5] Henderson P.J.F. Curr. Opin. Struct. Biol. 1:590-601(1991).[6] Culham D.E., Lasby B., Marangoni A.G., Milner J.L., Steer B.A., van Nues R.W., Wood J.M.-J. Mol. Biol. 229:268-276(1993).

633. Synaptobrevin signature

Synaptobrevin [1] is an intrinsic membrane protein of small synaptic vesicles whose function is not yet known, but which is highly conserved in mammals, electric ray (where its is known as VAMP-1), Drosophila and yeast [2]. In yeast there are two closely related forms of synaptobrevin (genes SNC1 andSNC2) while in mammals there is at least 4 (genes SYB1, SYB2, SYB3 and SYBL1).Structurally synaptobrevin consist of a N-terminal cytoplasmic domain of from 90 to 110 residues, followed by a transmembrane region, and then by a short (from 2 to 22 residues) C-terminal intravesicular domain. As a signature pattern for synaptobrevin, a highly conserved stretch of residues located in the central part of the sequence was selected.

Consensus pattern: N-[LIVM]-[DENS]-[KL]-V-x-[DEQ]-R-x(2)-[KR]-[LIVM]-[STDE]- x-[LIVM]-x-[DE]-[KR]-[TA]-[DE]

[1] Suedhof T.C., Baumert M., Perin M.S., Jahn R. Neuron 2:1475-1481(1989).[2] Gerst J.E., Rodgers L., Riggs M., Wigler M. Proc. Natl. Acad. Sci. U.S.A. 89:4338-4342(1992).

634. TBC domain. Identification of a TBC domain in GYP6_YEAST and GYP7_YEAST, which are GTPase activator proteins of yeast Ypt6 and Ypt7, imply that these domains are GTPase activator proteins of Rab-like small GTPases. Number of members: 55

[1] Medline: 96032578. Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16. Richardson PM, Zon LI; Oncogene 1995;11:1139-1148.

[2]Medline: 97398935. A shared domain between a spindle assembly checkpoint protein and Ypt/Rab-specific GTPase-activators. Neuwald AF; Trends Biochem Sci 1997;22:243-244.

635. Transcription factor TFIID repeat signature (TBP)

Transcription factor TFIID (or TATA-binding protein, TBP) [1,2] is a general factor that plays a major role in the activation of eukaryotic genes transcribed by RNA polymerase II. TFIID binds specifically to the TATA box promoter element which lies close to the position of transcription initiation. There is a remarkable degree of sequence conservation of a C-terminal domain of about 180 residues in TFIID from various eukaryotic sources. This region isnecessary and sufficient for TATA box binding. The most significant structural feature of this domain is the presence of two conserved repeats of a 77 amino-acid region. The intramolecular symmetry generates a saddle-shaped structure that sits astride the DNA [3]. Drosophila TRF (TBP-related factor) [4] is a sequence-specific transcription factor that also binds to the TATA box and is highly similar to TFIID. Archaebacteria also possess a TBP homolog [5]. A signature pattern that spans the last 50 residues of the repeated region has been derived.-

Consensus pattern: Y-x-P-x(2)-[IF]-x(2)-[LIVM](2)-x-[KRH]-x(3)-P-[RKQ]-x(3)- L-[LIVM]-F-x-[STN]-G-[KR]-[LIVM]-x(3)-G-[TAGL]-[KR]-x(7)- [AGC]-x(7)-[LIVM

[1] Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M., Roeder R.G. Nature 346:387-390(1990).[2] Gash A., Hoffmann A., Horikoshi M., Roeder R.G., Chua N.-H. Nature 346:390-394(1990).[3] Nikolov D.B., Hu S.-H., Lin J., Gasch A., Hoffmann A., Horikoshi M., Chua N.-H., Roeder R.G., Burley S.K. Nature 360:40-46(1992).[4] Crowley T.E., Hoey T., Liu J.-K., Jan Y.N., Jan L.Y., Tjian R. Nature 361:557-561(1993).[5] Marsh

T.L., Reich C.I., Whitelock R.B., Olsen G.J. Proc. Natl. Acad. Sci. U.S.A. 91:4180-4184(1994).

636. Translationally controlled tumor protein signatures (TCTP)

Mammalian translationally controlled tumor protein (TCTP) (or P23) is a protein which has been found to be preferentially synthesized in cells during the early growth phase of some types of tumor [1,2], but which is also expressed in normal cells. The physiological function of TCTP is still not known. It is a hydrophilic protein of 18 to 20 Kd. Close homologs have been found in plants [3], earthworm [4], Caenorhabditis elegans (F52H2.11), Hydra, budding yeast (YKL056c) [5] and fission yeast (SpAC1F12.02c) Two of the best conserved regions have been selected as signature patterns for TCTP.

Consensus pattern: [IFA]-[GA]-[GAS]-N-[PAK]-S-[GA]-E-[GDE]-[PAGE]-[DEQGA]

Consensus pattern: [FLVH]-[FY]-[IVCT]-G-E-x-[MA]-x(2,5)-[DEN]-[GAST]-x-[LV]-[AV]-x(3)-[FYW]

[1] Boehm H., Beendorf R., Gaestel M., Gross B., Nuernberg P., Kraft R., Otto A., Bielka H. Biochem. Int. 19:277-286(1989).[2] Makrides S., Chitpatima S.T., Bandyopadhyay R., Brawerman G. Nucleic Acids Res. 16:2350-2350(1988).[3] Pay A., Heberle-Bors E., Hirt H. Plant Mol. Biol. 19:501-503(1992).[4] Stuerzenbaum S.R., Kille P., Morgan A.J. Biochim. Biophys. Acta 1398:294-304(1998).[5] Rasmussen S.W. Yeast 10:S63-S68(1994).

637. TFIIS zinc ribbon domain signature

Transcription factor S-II (TFIIS) [1] is a eukaryotic protein necessary for efficient RNA polymerase II transcription elongation, past template-encoded pause sites. TFIIS shows DNA-binding activity only in the presence of RNA polymerase II. It is a protein of about 300 amino acids whose sequence is highly conserved in mammals, Drosophila, yeast (where it was first known as PPR2, a transcriptional regulator of URA4, and then as DST1, the DNA strand transfer protein alpha [2]) and in the archaebacteria Sulfolobus acidocaldarius [3].This family also includes the eukaryotic and archebacterial RNA polymerase subunits of the 15 Kd / M family (see <PDOC00790>) as well as the following viral proteins: - Vaccinia virus RNA polymerase 30 Kd subunit (rpo30) [4]. - African swine fever virus protein I243L [5].The best conserved region of all these proteins contains four cysteines that bind a zinc ion and fold in a conformation termed a 'zinc ribbon' [6]. Besides these cysteines, there are a

number of other conserved residues which can be used to help define a specific pattern for this type of domain.

Consensus pattern: C-x(2)-C-x(9)-[LIVMQSAR]-[QH]-[STQL]-[RA]-[SACR]-x-[DE]-[DET]-[PGSEA]-x(6)-C-x(2,5)-C-x(3)-[FW] [The four C's are zinc ligands]

[1] Hirashima S., Hirai H., Nakanishi Y., Natori S. J. Biol. Chem. 263:3858-3863(1988).[2] Kipling D., Kearsey S.E. Nature 353:509-509(1991).[3] Langer D., Zillig W. Nucleic Acids Res. 21:2251-2251(1993).[4] Ahn B.-Y., Gershon P.D., Jones E.V., Moss B. Mol. Cell. Biol. 10:5433-5441(1990).[5] Rodriguez J.M., Salas M.L., Vinuela E. Virology 186:40-52(1992).[6] Qian X., Jeon C., Yoon H., Agarwal K., Weiss M.A. Nature 365:277-279(1993).

638. Tetrahydrofolate dehydrogenase/cyclohydrolase signatures (THF DHG CYH)

Enzymes that participate in the transfer of one-carbon units are involved in various biosynthetic pathways. In many of these processes the transfers of one-carbon units are mediated by the coenzyme tetrahydrofolate (THF). Various reactions generate one-carbon derivatives of THF which can be interconverted between different oxidation states by formyltetrahydrofolate synthetase(EC 6.3.4.3), methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5 or EC 1.5.1.15) and methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9).The dehydrogenase and cyclohydrolase activities are expressed by a variety of multifunctional enzymes: - Eukaryotic C-1-tetrahydrofolate synthase (C1-THF synthase), which catalyzes all three reactions described above. Two forms of C1-THF synthases are known [1], one is located in the mitochondrial matrix, while the second one is cytoplasmic. In both forms the dehydrogenase/cyclohydrolase domain is located in the N-terminal section of the 900 amino acids protein and consists of about 300 amino acid residues. The C1-THF synthases are NADP- dependent. - Eukaryotic mitochondrial bifunctional dehydrogenase/cyclohydrolase [2]. This is an homodimeric NAD-dependent enzyme of about 300 amino acid residues. - Bacterial folD [3]. FolD is an homodimeric bifunctional NADP-dependent enzyme of about 290 amino acid residues. The sequence of the dehydrogenase/cyclohydrolase domain is highly conserved in all forms of the enzyme. Two conserved regions have been selected as signature patterns. The first one is located in the N-terminal part of these enzymes and contains three acidic residues. The second pattern is a highly conserved sequence of 9 amino acids which is located in the C-terminal section.

Consensus pattern: [EQ]-x-[EQK]-[LIVM](2)-x(2)-[LIVM]-x(2)-[LIVMY]-N-x-[DN]- x(5)-[LIVMF](3)-Q-L-P-[LV]

Consensus pattern: P-G-G-V-G-P-[MF]-T-[IV]

[1] Shannon K.W., Rabinowitz J.C. J. Biol. Chem. 263:7717-7725(1988).[2] Belanger C., Mackenzie R.E. J. Biol. Chem. 264:4837-4843(1989).[3] d'Ari L., Rabinowitz J.C. J. Biol. Chem. 266:23953-23958(1991).

639. Triosephosphate isomerase active site (TIM)

Triosephosphate isomerase (EC 5.3.1.1) (TIM) [1] is the glycolytic enzyme that catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. TIM plays an important role in several metabolic pathways and is essential for efficient energy production. It is a dimer of identical subunits, each of which is made up of about 250 amino-acid residues. A glutamic acid residue is involved in the catalytic mechanism [2]. The sequence around the active site residue is perfectly conserved in all known TIM's and can be used as a signature pattern for this type of enzyme.

Consensus pattern: [AV]-Y-E-P-[LIVM]-W-[SA]-I-G-T-[GK] [E is the active site residue]

[1] Lolis E., Alber T., Davenport R.C., Rose D., Hartman F.C., Petsko G.A. Biochemistry 29:6609-6618(1990).[2] Knowles J.R. Nature 350:121-124(1991).

640. Thymidine kinase cellular-type signature (TK)

Thymidine kinase (TK) (EC 2.7.1.21) is an ubiquitous enzyme that catalyzes the ATP-dependent phosphorylation of thymidine. A comparison of TK sequences has shown [1,2,3] that there are two different families of TK. One family groups together TK from herpes viruses as well as cellular thymidylate kinases, while the second family currently consists of TK from the following sources: - Vertebrates. - Bacterial. - Bacteriophage T4. – Pox viruses. - African swine fever virus (ASF). - Fish lymphocystis disease virus (FLDV).A conserved region which is located in the C-terminal section of these enzymes has been selected as a signature pattern for this family of TKA.

Consensus pattern: [GA]-x(1,2)-[DE]-x-Y-x-[STAP]-x-C-[NKR]-x-[CH]-[LIVMFYWH]

[1] Boyle D.B., Coupar B.E.H., Gibbs A.J., Seigman L.J., Both G.W. Virology 156:355-365(1987).[2] Blasco R., Lopez-Otin C., Munoz M., Bockamp E.-O., Simon-Mateo C., Vinuela E. Virology 178:301-304(1990).[3] Robertson G.R., Whalley J.M. Nucleic Acids Res. 16:11303-11317(1988).

641. Thymidine kinase from herpesvirus (TK herpes)

[1]

Medline: 96003730

Crystal structures of the thymidine kinase from herpes simplex virus type-1 in complex with deoxythymidine and ganciclovir.

Brown DG, Visse R, Sandhu G, Davies A, Rizkallah PJ, Melitz C, Summers WC, Sanderson MR;

Nat Struct Biol 1995;2:876-881.

Number of members: 65

642. Nuclear transition protein 2 signatures (TP2)

In mammals, the second stage of spermatogenesis is characterized by the conversion of nucleosomal chromatin to the compact, non-nucleosomal and transcriptionally inactive form found in the sperm nucleus. This condensation is associated with a double-protein transition. The first transition corresponds to the replacement of histones by several spermatid-specific proteins, also called transition proteins, which are themselves replaced by protamines during the second transition. Nuclear transition protein 2 (TP2) is one of those spermatid-specific proteins. TP2 is a basic, zinc-binding protein [1] of 116 to 137 amino-acid residues. Structurally, TP2 consists of three distinct parts: a conserved serine-rich N-terminal domain of about 25 residues, a variable central domain of 20 to 50 residues which contains cysteine residues, and a conserved C-terminal domain of about 70 residues rich in lysines and arginines. Two signature patterns for TP2 have been developed: one located in the N-terminal domain, the other in the C-terminal.

Consensus pattern: H-x(3)-H-S-[NS]-S-x-P-Q-S

Consensus pattern: K-x-R-K-x(2)-E-G-K-x(2)-K-[KR]-K

[1] Baskaran R., Rao M.R.S. Biochem. Biophys. Res. Commun. 179:1491-1499(1991).

643. Thiamine pyrophosphate enzymes signature (TTP enzymes)

A number of enzymes require thiamine pyrophosphate (TPP) (vitamin B1) as a cofactor. It has been shown [1] that some of these enzymes are structurally related. These related TPP

enzymes are: - Pyruvate oxidase (POX) (EC 1.2.3.3) Reaction catalyzed: pyruvate + orthophosphate + O(2) + H(2)O = acetyl phosphate + CO(2) + H(2)O(2). - Pyruvate decarboxylase (PDC) (EC 4.1.1.1) Reaction catalyzed: pyruvate = acetaldehyde + CO(2). - Indolepyruvate decarboxylase (EC 4.1.1.74) [2] Reaction catalyzed: indole-3-pyruvate = indole-3-acetaldehyde + CO(2). - Acetolactate synthase (ALS) (EC 4.1.3.18) Reaction catalyzed: 2 pyruvate = acetolactate + CO(2). - Benzoylformate decarboxylase (BFD) (EC 4.1.1.7) [3] Reaction catalyzed: benzoylformate = benzaldehyde + CO(2). A conserved region which is located in their C-terminal section has been selected as a signature pattern for these enzymes.

Consensus pattern: [LIVMF]-[GSA]-x(5)-P-x(4)-[LIVMFYW]-x-[LIVMF]-x-G-D-[GSA]-[GSAC]

[1] Green J.B.A. FEBS Lett. 246:1-5(1989).[2] Koga J., Adachi T., Hidaka H. Mol. Gen. Genet. 226:10-16(1991).[3] Tsou A.Y., Ransom S.C., Gerlt J.A., Buechter D.D., Babbitt P.C., Kenyon G.L. Biochemistry 29:9856-9862(1990).

644. TPR Domain

[1]

Medline: 95397415

Tetratrico peptide repeat interactions: to TPR or not to TPR?

Lamb JR, Tugendreich S, Hieter P;

Trends Biochem Sci 1995;20:257-259.

[2]Medline: 98151343

The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.

Das AK, Cohen PW, Barford D;

EMBO J 1998;17:1192-1199.

Number of members: 621

645. Uroporphyrin-III C-methyltransferase signatures (TP methylase)

Uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (SUMT) [1,2] catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to the C-2 and C-7atoms of

uroporphyrinogen III to yield precorrin-2 via the intermediate formation of precorrin-1. SUMT is the first enzyme specific to the cobalamin pathway and precorrin-2 is a common intermediate in the biosynthesis of corrinoids such as vitamin B12, siroheme and coenzyme F430.The sequences of SUMT from a variety of eubacterial and archaebacterial species are currently available. In species such as Bacillus megaterium (gene cobA), Pseudomonas denitrificans (cobA) or Methanobacterium ivanovii (gene corA) SUMT is a protein of about 25 to 30 Kd. In Escherichia coli and related bacteria, the cysG protein, which is involved in the biosynthesis of siroheme, is a multifunctional protein composed of a N-terminal domain, probably involved in transforming precorrin-2 into siroheme, and a C-terminal domain which has SUMT activity. The sequence of SUMT is related to that of a number of P. denitrificans and Salmonella typhimurium enzymes involved in the biosynthesis of cobalamin which also seem to be SAM-dependent methyltransferases [3,4]. The similarity is especially strong with two of these enzymes: cobI/cbiL which encodes S-adenosyl-L-methionine--precorrin-2 methyltransferase and cobM/cbiF whose exact function is not known. Two signature patterns have been developed for these enzymes. The first corresponds to a well conserved region in the N-terminal extremity (called region 1 in [1,3]) and the second to a less conserved region located in the central part of these proteins (this pattern spans what are called regions 2 and 3 in [1,3]).

Consensus pattern: [LIVM]-[GS]-[STAL]-G-P-G-x(3)-[LIVMFY]-[LIVM]-T-[LIVM]-[KRHQG]-[AG]

Consensus pattern: V-x(2)-[LI]-x(2)-G-D-x(3)-[FYW]-[GS]-x(8)-[LIVF]-x(5,6)-[LIVMFYWPAC]-x-[LIVMY]-x-P-G

[1] Blanche F., Robin C., Couder M., Faucher D., Cauchois L., Cameron B., Crouzet J. J. Bacteriol. 173:4637-4645(1991).[2] Robin C., Blanche F., Cauchois L., Cameron B., Couder M., Crouzet J. J. Bacteriol. 173:4893-4896(1991).[3] Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C., Blanche F., Thibaut D., Debussche L. J. Bacteriol. 172:5980-5990(1990).[4] Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S., Church G.M. J. Bacteriol. 175:3303-3316(1993).[5] Mattheakis L.C., Shen W.H., Collier R.J. Mol. Cell. Biol. 12:4026-4037(1992).

646. Tudor domain

Domain of unknown function present in several RNA-binding proteins. copies in the Drosophila Tudor protein. Slight ambiguities in the alignment.Number of members: 18

[1]Medline: 97200561 Tudor domains in proteins that interact with RNA. Ponting CP; Trends Biochem Sci 1997;22:51-52. [2]Medline: 97157029 The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in Drosophila melanogaster development. Callebaut I, Mornon JP; Biochem J 1997;321:125-132.

647. Terpene synthase family

It has been suggested that this gene family be designated tps (for terpene synthase) [1]. It has been split into six subgroups on the basis of phylogeny, called tpsa-tpsf.

tpsa includes vetispiridiene synthase Swiss:Q39979, 5-epi-aristolochene synthase, Swiss:Q40577 and (+)-delta-cadinene synthase Swiss:P93665.

tpsb includes (-)-limonene synthase, Swiss:Q40322.

tpsc includes kaurene synthase A, Swiss:O04408.

tpsd includes taxadiene synthase, Swiss:Q41594, pinene synthase, Swiss:O24475 and myrcene synthase, Swiss:O24474.

tpse includes kaurene synthase B.

tpsf includes linalool synthase.

Number of members: 51

[1]

Medline: 97413772

Monoterpene synthases from grand fir (Abies grandis). cDNA isolation, characterization, and functional expression of myrcene synthase, (-)-(4S)-limonene synthase, and (-)-(1S,5S)-pinene synthase.

Bohlmann J, Steele CL, Croteau R;

J Biol Chem 1997;272:21784-21792.

648. ThiF family

This family contains a repeated domain in ubiquitin activating enzyme E1 and members of the bacterial

ThiF/MoeB/HesA family.Number of members: 87

649. Thioester dehydrase

Members of this family are involved in fatty acid biosynthesis.

Number of members: 19

[1]

Medline: 96398612

Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site.

Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL;

Structure 1996;4:253-264.

Database Reference: SCOP; 1mka; fa; [SCOP-USA][CATH-PDBSUM]

Database reference: PFAMB; PB058036;

650. Tub family signatures

The mouse tubby mutation is the cause of maturity-onset obesity, insulin resistance and sensory deficits. This mutation maps to a gene, tub [1,2],which codes for a protein that belongs to a family which currently consists of the following members: - Mammalian tub, an hydrophilic protein of about 500 residues, which could be involved in the hypothalamic regulation of body weight. - Human protein TULP1 [3] which may be involved in retinis pigmentosa 14, a retinal degeneration disease. - Mouse protein p4-6 whose function is not known. - Caenorhabditis elegans hypothetical protein F10B5.4. - Several fragmentary sequences from plants, Drosophila and human ESTs. While the N-terminal part of these protein is not conserved in length nor in the sequence, the C-terminal 250 residues are highly conserved. Therefore, two regions were selected in the C-terminal part as signature patterns. The secondr egion is located at the C-terminal extremity and contains a penultimate cysteine residue that could be critical to the normal functioning of these proteins.

Consensus pattern: F-[KHQ]-G-R-V-[ST]-x-A-S-V-K-N-F-Q

Consensus pattern: A-F-[AG]-I-[SAC]-[LIVM]-[ST]-S-F-x-[GST]-K-x-A-C-E

[1] Kleyn P.W., Fan W., Kovats S.G., Lee J.L., Pulido J.C., Wu Y., Berkemeier L.R., Misumi D.J., Holmgren L., Charlat O., Woolf E.A., Tayber O., Brody T., Shu P., Hawkins F.,

Kennedy B., Baldini L., Ebeling C., Alperin G.D., Deeds J., Lakey N.D., Culpepper J., Chen H., Gluecksmann-Kuis M.A., Carlson G.A., Duyk G.M., Moore K.J. Cell 85:281-290(1996).[2] Noben-Trauth K., Naggert J.K., North M.A., Nishina P.M. Nature 380:534-538(1996).[3] North M.A., Naggert J.K., Yan Y., Noben-Trauth K., Nishina P.M. Proc. Natl. Acad. Sci. U.S.A. 94:3128-3133(1997).

651. Eukaryotic DNA topoisomerase I active site

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type Itopoisomerases act by catalyzing the transient breakage of DNA, one strand at a time, and the subsequent rejoining of the strands. When a eukaryotic type 1topoisomerase breaks a DNA backbone bond, it simultaneously forms a protein-DNA link where the hydroxyl group of a tyrosine residue is joined to a 3'-phosphate on DNA, at one end of the enzyme-severed DNA strand. In eukaryotes and pox virus topoisomerases I, there are a number of conserved residues in the region around the active site tyrosine.

Consensus pattern: [DEN]-x(6)-[GS]-[IT]-S-K-x(2)-Y-[LIVM]-x(3)-[LIVM] [Y is the active site tyrosine]

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).[2] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).[3] Lynn R.M., Bjornsti M.-A., Caron P.R., Wang J.C. Proc. Natl. Acad. Sci. U.S.A. 86:3559-3563(1989).[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).[E1]

652. Transaldolase signatures

Transaldolase (EC 2.2.1.2) catalyzes the reversible transfer of a three-carbonketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form erythrose 4-phosphate and fructose 6-phosphate. This enzyme, together with transketolase, provides a link between the glycolytic and pentose-phosphate pathways. Transaldolase is an enzyme of about 34 Kd whose sequence has been well conserved throughout evolution. A lysine has been implicated [1]in the catalytic mechanism of the enzyme; it acts as a nucleophilic group that attacks the carbonyl group of fructose-6-phosphate.Transaldolase is evolutionary related [2] to a bacterial protein of about 20Kd (known as talC in Escherichia coli), whose exact function is not yet known. Two signature patterns have been developed for these proteins. The first,

located in the N-terminal section, contains a perfectly conserved pentapeptide; these cond, includes the active site lysine.

Consensus pattern: [DG]-[IVSA]-T-[ST]-N-P-[STA]-[LIVMF](2)

Consensus pattern: [LIVM]-x-[LIVM]-K-[LIVM]-[PAS]-x-[ST]-x-[DENQPAS]-G- [LIVM]-x-[AGV]-x-[QEKRST]-x-[LIVM] [K is the active site residue]

[1] Miosga T., Schaaff-Gerstenschlaeger I., Franken E., Zimmermann F.K. Yeast 9:1241-1249(1993).[2] Reizer J., Reizer A., Saier M.H. Jr. Microbiology 141:961-971(1995).

653. (Transpeptidase) Penicillin binding protein transpeptidase domain

The active site serine (residue 337 in Swiss:P14677) is conserved in all members of this family.

[1] Pares S, Mouz N, Petillot Y, Hakenbeck R, Dideberg O Nat Struct Biol 1996;3:284-289.

654. Trehalase signatures

Trehalase (EC 3.2.1.28) is the enzyme responsible for the degradation of the disaccharide alpha, alpha-trehalose yielding two glucose subunits [1]. It is an enzyme found in a wide variety of organisms and whose sequence has been highly conserved throughout evolution. Two of the most highly conserved regions have been selected as signature patterns. The first pattern is located in the central section, the second one is in the C-terminal region.

Consensus pattern: P-G-G-R-F-x-E-x-Y-x-W-D-x-Y

Consensus pattern: Q-W-D-x-P-x-[GA]-W-[PAS]-P

[1] Kopp M., Mueller H., Holzer H. J. Biol. Chem. 268:4766-4774(1993).[2] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993).[E1]

655. Trehalose-6-phosphate synthase domain

OtsA (Trehalose-6-phosphate synthase) is homologous to regions in the subunits of yeast trehalose-6-phosphate synthase/phosphate complex, [1].

[1] Kaasen I, McDougall J, Strom AR; Gene 1994;145:9-15.

656. Tropomyosins signature

Tropomyosins [1,2] are family of closely related proteins present in muscle and non-muscle cells. In striated muscle, tropomyosin mediate the interactions between the troponin complex and actin so as to regulate muscle contraction. The role of tropomyosin in smooth muscle and non-muscle tissues is not clear. Tropomyosin is an alpha-helical protein that forms a coiled-coil dimer. Muscle isoforms of tropomyosin are characterized by having 284 amino acid residues and a highly conserved N-terminal region, whereas non-muscle forms are generally smaller and are heterogeneous in their N-terminal region. The signature pattern for tropomyosins is based on a very conserved region in the C-terminal section of tropomyosins and which is present in both muscle and non-muscle forms.

Consensus pattern: L-K-E-A-E-x-R-A-E

[1] Smilie L.B. Trends Biochem. Sci. 4:151-155(1979).[2] McLeod A.R. BioEssays 6:208-212(1986).

657. Troponin

Troponin (Tn) contains three subunits, Ca2+ binding (TnC), inhibitory (TnI), and tropomyosin binding (TnT). this Pfam contains members of the TnT subunit.

Troponin is a complex of three proteins, Ca2+ binding (TnC), inhibitory (TnI), and tropomyosin binding (TnT).

The troponin complex regulates Ca++ induced muscle contraction.

This family includes troponin T and troponin I. Troponin I binds to actin and troponin T binds to tropomyosin.

Number of members: 81 [1]

Medline: 87144593

Structure of co-crystals of tropomyosin and troponin.

White SP, Cohen C, Phillips GN Jr;

Nature 1987;325:826-828. [2]

Medline: 95155315

A direct regulatory role for troponin T and a dual role for troponin C in the Ca2+ regulation of muscle contraction.

Potter JD, Sheng Z, Pan BS, Zhao J;

J Biol Chem 1995;270:2557-2562.

[3]Medline: 95324796

The troponin complex and regulation of muscle contraction.

Farah CS, Reinach FC;

FASEB J 1995;9:755-767.

658. (Tryp mucin) Mucin-like glycoprotein

This family of trypanosomal proteins resemble vertebrate mucins. The protein consists of three regions. The N and C terminii are conserved between all members of the family, whereas the central region is not well conserved and contains a large number of threonine residues which can be glycosylated [1].

Indirect evidence suggested that these genes might encode the core protein of parasite mucins, glycoproteins that were proposed to be involved in the interaction with, and invasion of, mammalian host cells.

[1] Di Noia JM, Sanchez DO, Frasch AC; J Biol Chem 1995;270:24146-24149.

[2] Di Noia JM, D'Orso I, Aslund L, Sanchez DO, Frasch AC; J Biol Chem 1998;273:10843-10850.

659. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each differentamino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse interms of subunit size and of quaternary structure. A few years ago it was found [2] that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific for

arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiarystructure based on a Rossmann fold. Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

660. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1b)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. A few years ago it was found [2] that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific forarginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiary structure based on a Rossmann fold. Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

661. (tRNA-synt 1C) tRNA synthetases class I (E and Q)

Other tRNA synthetase sub-families are too dissimilar to be included.
This family includes only glutamyl and glutaminyl tRNA synthetases.
In some organisms, a single glutamyl-tRNA synthetase aminoacylates both tRNA(Glu) and tRNA(Gln).

[1] Rath VL, Silvian LF, Beijer B, Sproat BS, Steitz TA; Structure 1998;6:439-449.

662. (tRNA-synt 1d) tRNA synthetases class I (R)

Other tRNA synthetase sub-families are too dissimilar to be included.
This family includes only arginyl tRNA synthetase.

663. Aminoacyl-transfer RNA synthetases class-II signatures (tRNA synt 2)
Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse interms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.
Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Delarue M., Moras D. BioEssays 15:675-687(1993).[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991). [5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).[6] Cusack S. Biochimie 75:1077-1081(1993).[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

664. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1e)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. A few years ago it was found [2] that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific forarginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiary structure based on a Rossmann fold.

Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

665. Aminoacyl-transfer RNA synthetases class-II signatures (tRNA synt 2b)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse interms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Delarue M., Moras D. BioEssays 15:675-687(1993).[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991). [5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).[6] Cusack S. Biochimie 75:1077-1081(1993).[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

666. Thaumatin family signature

Thaumatin [1] is an intensively sweet-tasting protein (100 000 times sweeter than sucrose on a molar basis) from Thaumatococcus daniellii, an African brush. The protein is made of about 200 residues and contains 8 disulfide bonds. A number of proteins have been found to be related to thaumatins. These protein are listed below (references are only provided for recently determined sequences). - A maize alpha-amylase/trypsin inhibitor. - Two tobacco pathogenesis-related proteins: PR-R major and minor forms, which are induced after

infection with viruses. - Salt-induced protein NP24 from tomato. - Osmotin, a salt-induced protein from tobacco. - Osmotin-like proteins OSML13, OSML15 and OSML81 from potato [2]. - P21, a leaf protein from soybean. - PWIR2, a leaf protein from wheat. - Zeamatin, a maize antifunal protein [3].The exact biological function of all these proteins is not yet known. A conserved region that includes three cysteine residues known (in thaumatin) to be involved in disulfide bonds has been selected as a signature pattern.

+--+ | +-----------------+ | | ******* | | | xxCxxxxxxxxxxxxxxxxCxxCxxCxCxxxxxxxxxxxxxxCxxCxCxxxCxCxxCCxCxxxCxxxxxC xxxCx | | | | | | | | || | | | +--+ +-+ | +---+ +--++-+ | +--------------------------+'C': conserved cysteine involved in a disulfide bond.'*': position of the pattern.

Consensus pattern: G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-G-x(2,3)-C

[1] Edens L., Heslinga L., Klok R., Ledeboer A.M., Maat J., Toonen M.Y., Visser C., Verrips C.T. Gene 18:1-12(1982).[2] Zhu B., Chen T.H.H., Li P.H. Plant Physiol. 108:929-937(1995).[3] Malehorn D.E., Borgmeyer J.R., Smith C.E., Shah D.M.; Plant Physiol. 106:1471-1481(1994).

667. Thiolases signatures

Two different types of thiolase [1,2,3] are found both in eukaryotes and in prokaryotes: acetoacetyl-CoA thiolase (EC 2.3.1.9) and 3-ketoacyl-CoA thiolase(EC 2.3.1.16). 3-ketoacyl-CoA thiolase (also called thiolase I) has a broad chain-length specificity for its substrates and is involved in degradative pathways such as fatty acid beta-oxidation. Acetoacetyl-CoA thiolase (also called thiolase II) is specific for the thiolysis of acetoacetyl-CoA and involved in biosynthetic pathways such as poly beta-hydroxybutyrate synthesisor steroid biogenesis. In eukaryotes, there are two forms of 3-ketoacyl-CoA thiolase: one located in the mitochondrion and the other in peroxisomes. There are two conserved cysteine residues important for thiolase activity. The first located in the N-terminal section of the enzymes is involved in the formation of an acyl-enzyme intermediate; the second located at the C-terminal extremity is the active site base involved in deprotonation in the condensation reaction. Mammalian nonspecific lipid-transfer protein (nsL-TP) (also known as sterol carrier protein 2) is a protein which seems to exist in two different forms: a 14 Kd protein (SCP-2) and a larger 58 Kd protein (SCP-x). The former is found in the cytoplasm or the mitochondria and is involved in lipid transport; the latter is found in peroxisomes. The C-terminal part of SCP-x is identical to

SCP-2 while the N-terminal portion is evolutionary related to thiolases[4]. Three signature patterns have been developed for this family of proteins, two of which are based on the regions around the biologically important cysteines. The third is based on a highly conserved region in the C-terminal part of these proteins.

Consensus pattern: [LIVM]-[NST]-x(2)-C-[SAGLI]-[ST]-[SAG]-[LIVMFYNS]-x- [STAG]-[LIVM]-x(6)-[LIVM] [C is involved in formation of acyl-enzyme intermediate]

Consensus pattern: N-x(2)-G-G-x-[LIVM]-[SA]-x-G-H-P-x-[GA]-x-[ST]-G

Consensus pattern: [AG]-[LIVMA]-[STAGCLIVM]-[STAG]-[LIVMA]-C-x-[AG]-x-[AG]-x- [AG]-x-[SAG] [C is the active site residue]

[1] Peoples O.P., Sinskey A.J. J. Biol. Chem. 264:15293-15297(1989).[2] Yang S.-Y., Yang X.-Y.H., Healy-Louie G., Schulz H., Elzinga M. J. Biol. Chem. 265:10424-10429(1990).[3] Igual J.C., Gonzalez-Bosch C., Dopazo J., Perez-Ortin J.E. J. Mol. Evol. 35:147-155(1992).[4] Baker M.E., Billheimer J.T., Strauss J.F. III DNA Cell Biol. 10:695-698(1991).

668. Thioredoxin family active site

Thioredoxins [1 to 4] are small proteins of approximately one hundred amino-acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is wellconserved. Bacteriophage T4 also encodes for a thioredoxin but its primary structure is not homologous to bacterial, plant and vertebrate thioredoxins. A number of eukaryotic proteins contain domains evolutionary related tothioredoxin, all of them seem to be protein disulphide isomerases (PDI). PDI(EC 5.3.4.1) [5,6,7] is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: - PDI major isozyme; a multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deiodinase (EC 3.8. 1.4), as glutathione-insulin transhydrogenase (EC 1.8.4.2) and as a thyroid hormone-binding protein ! - ERp60 (ER-60; 58 Kd microsomal protein). ERp60 was originally thought to be a phosphoinositide-specific phospholipase C isozyme and later to be a protease. - ERp72. - P5.All PDI contains two or three (ERp72) copies of the thioredoxin domain. Bacterial proteins that act as thiol:disulfide interchange proteins thatallows disulfide bond formation in

some periplasmic proteins also contain a thioredoxin domain. These proteins are: - Escherichia coli dsbA (or prfA) and its orthologs in Vibrio cholerae (tcpG) and Haemophilus influenzae (por). - Escherichia coli dsbC (or xpRA) and its orthologs in Erwinia chrysanthemi and Haemophilus influenzae. - Escherichia coli dsbD (or dipZ) and its Haemophilus influenzae ortholog. - Escherichia coli dsbE (or ccmG) and orthologs in Haemophilus influenzae, Rhodobacter capsulatus (helX), Rhiziobiacae (cycY and tlpA).

Consensus pattern: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C- [GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT] [The two C's form the redox-active bond]

[1] Holmgren A. Annu. Rev. Biochem. 54:237-271(1985).[2] Gleason F.K., Holmgren A. FEMS Microbiol. Rev. 54:271-297(1988).[3] Holmgren A. J. Biol. Chem. 264:13963-13966(1989).[4] Eklund H., Gleason F.K., Holmgren A. Proteins 11:13-28(1991).[5] Freedman R.B., Hawkins H.C., Murant S.J., Reid L. Biochem. Soc. Trans. 16:96-99(1988).[6] Kivirikko K.I., Myllyla R., Pihlajaniemi T. FASEB J. 3:1609-1617(1989).[7] Freedman R.B., Hirst T.R., Tuite M.F. Trends Biochem. Sci. 19:331-336(1994).

669. (Transcript fac2) Transcription factor TFIIB repeat signature

In eukaryotes the initiation of transcription of protein encoding genes by polymerase II is modulated by general and specific transcription factors. The general transcription factors operate through common promoters elements (such as the TATA box). At least seven different proteins associates to form the general transcription factors: TFIIA, -IIB, -IID, -IIE, -IIF, -IIG, and -IIH[1].Transcription factor IIB (TFIIB) plays a central role in the transcription of class II genes, it associates with a complex of TFIID-IIA bound to DNA (DA complex) to form a ternary complex TFIID-IIA-IBB (DAB complex) which is then recognized by RNA polymerase II [2,3]. TFIIB is a protein of about 315 to 340amino acid residues which contains, in its C-terminal part an imperfect repeat of a domain of about 75 residues. This repeat could contribute an element of symmetry to the folded protein. The following proteins have been shown to be evolutionary related to TFIIB: - An archaebacterial TFIIB homolog. In Pyrococcus woesei a previously undetected open reading frame has been shown [4] to be highly related to TFIIB. - Fungal transcription factor IIIB 70 Kd subunit (gene PCF4/TDS4/BRF1) [5]. This protein is a general activator of RNA polymerase III transcription and plays a role analogous to that of TFIIB in pol III transcription. The central

section of the repeated domain, which is the most conserved part of that domain has been selected as a signature pattern.

Consensus pattern: G-[KR]-x(3)-[STAGN]-x-[LIVMYA]-[GSTA](2)-[CSAV]-[LIVM]-[LIVMFY]-[LIVMA]-[GSA]-[STAC

[1] Weinmann R. Gene Expr. 2:81-91(1992).[2] Hawley D. Trends Biochem. Sci. 16:317-318(1991).[3] Ha I., Lane W.S., Reinberg D. Nature 352:689-695(1991).[4] Ouzounis C., Sander C. Cell 71:189-190(1992).[5] Khoo B., Brophy B., Jackson S.P. Genes Dev. 8:2879-2890(1994).

670. (transcritp fact) MADS-box domain signature and profile

A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below: - Serum response factor (SRF) [1], a mammalian transcription factor that binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the transcription initiation site of genes such as c-fos. - Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. - Drosophila myocyte-specific enhancer factor 2 (MEF2). - Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes. - Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80). - Yeast transcription factor RLM1. - Yeast transcription factor SMP1. - Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower. - Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and Pistillata (PI) which act locally to specify the identity of the floral meristem and to determine sepal and petal development [4]. - Antirrhinum majus and tobacco homeotic protein deficiens (DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO cause the transformation of petals into sepals and of stamina into carpels. - Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6]. - Antirrhinum majus morphogenetic protein DEF H33 (squamosa).In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. A pattern that spans the complete length of the domain has been derived. The profile also spans the length of the MADS-box.

Consensus pattern: R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)- x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x- [LIVM](3)-x(6)-[LIVMF]-x(2)-[FY]

[1] Norman C., Runswick M., Pollock R., Treisman R. Cell 55:989-1003(1988).[2] Passmore S., Maine G.T., Elble R., Christ C., Tye B.-K. J. Mol. Biol. 204:593-606(1988).[3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. Nature 346:35-39(1990).[4] Goto K., Meyerowitz E.M. Genes Dev. 8:1548-1560(1994).[5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W.-E., Saedler H., Sommer H., Schwartz-Sommer Z. EMBO J. 11:4693-4704(1992).[6] Ma H., Yanofsky M.F., Meyerowitz E.M. Genes Dev. 5:484-495(1991).[E1]

671. Transketolase signatures

Transketolase (EC 2.2.1.1) (TK) catalyzes the reversible transfer of a two-carbon ketol unit from xylulose 5-phosphate to an aldose receptor, such as ribose 5-phosphate, to form sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate. This enzyme, together with transaldolase, provides a link between the glycolytic and pentose-phosphate pathways. TK requires thiamin pyrophosphate as a cofactor. In most sources where TK has been purified, it is a homodimer of approximately 70 Kd subunits. TK sequences from a variety of eukaryotic and prokaryotic sources [1,2] show that the enzyme has been evolutionarily conserved. In the peroxisomes of methylotrophic yeast Hansenula polymorpha, there is a highly related enzyme, dihydroxy-acetone synthase (DHAS) (EC 2.2.1.3) (also known as formaldehyde transketolase), which exhibits a very unusual specificity by including formaldehyde amongst its substrates. 1-deoxyxylulose-5-phosphate synthase (DXP synthase) [3] is an enzyme so far found in bacteria (gene dxs) and plants (gene CLA1) which catalyzes the thiamin pyrophosphoate-dependent acyloin condensation reaction between carbon atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D- xylulose-5-phosphate (dxp), a precursor in the biosynthetic pathway to isoprenoids, thiamin (vitamin B1), and pyridoxol (vitamin B6). DXP synthase is evolutionary related to TK. Two regions of TK have been selected as signature patterns. The first, located in the N-terminal section, contains a histidine residue which appears to function inproton transfer during catalysis [4]. The second, located in the central section, contains conserved acidic residues that are part of the active cleft and may participate in substrate-binding [4].

Consensus pattern: R-x(3)-[LIVMTA]-[DENQSTHKF]-x(5,6)-[GSN]-G-H-[PLIVMF]-[GSTA]-x(2)-[LIMC]-[GS

Consensus pattern: G-[DEQGSA]-[DN]-G-[PAEQ]-[ST]-[HQ]-x-[PAGM]-[LIVMYAC]-[DEFYW]-x(2)-[STAP]-x(2)-[RGA]

[1] Abedinia M., Layfield R., Jones S.M., Nixon P.F., Mattick J.S. Biochem. Biophys. Res. Commun. 183:1159-1166(1992).[2] Fletcher T.S., Kwee I.L., Nakada T., Largman C., Martin B.M. Biochemistry 31:1892-1896(1992).[3] Sprenger G.A., Schorken U., Wiegert T., Grolle S., De Graaf A.A., Taylor S.V., Begley T.P., Bringer-Meyer S., Sahm H. Proc. Natl. Acad. Sci. U.S.A. 94:12857-12862(1997).[4] Lindqvist Y., Schneider G., Ermler U., Sundstroem M. EMBO J. 11:2373-2379(1992).

672. Transmembrane 4 family signature

Recently a number of eukaryotic cell surface antigens have been found to be evolutionary related [1,2,3]. The proteins known to belong to this family are listed below: - Mammalian antigen CD9 (MIC3); A protein involved in platelet activation and aggregation. - Mammalian leukocyte antigen CD37, expressed on B lymphocytes. - Mammalian leukocyte antigen CD53 (OX-44), which may be involved in growth regulation in hematopoietic cells. - Mammalian lysosomal membrane protein CD63 (melanoma-associated antigen ME491; antigen AD1). - Mammalian antigen CD81 (cell surface protein TAPA-1), which may play an important role in the regulation of lymphoma cell growth. - Mammalian antigen CD82 (protein R2; antigen C33; Kangai 1 (KAI1)), which associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway. - Mammalian antigen CD151 (SFA-1; platelet-endothelial tetraspan antigen 3 (PETA-3)). - Mammalian cell surface glycoprotein A15 (TALLA-1; MXS1). - Mammalian novel antigen 2 (NAG-2). - Human tumor-associated antigen CO-029. - Schistosoma mansoni and japonicum 23 Kd surface antigen (SM23 / SJ23).These proteins share the following characteristics: they all seem to be type III membrane proteins (type III proteins are integral membrane proteins that contain a N-terminal membrane-anchoring domain which is not cleaved during biosynthesis and which functions both as a translocation signal and as a membrane anchor); they also contain three additional transmembrane regions, at least seven conserved cysteines residues, and are of approximately the same size (218 to 284 residues). These proteins are collectively know as the 'transmembrane 4 super family' (TM4) because they span the plasma membrane four times. A schematic diagram of the domain structure of these proteins isshown below. +-+-----+-------+----+-----+-----+------------

----------------+-----+----+ | | TMa | Extra | TM2| Cyt | TM3 | Extracellular | TM4 | Cyt| +-+-----+-------+----C-----C-----+--------CC---------C-----C---+-----C----+ *********Cyt : cytoplasmic domain. TMa : transmembrane anchor.TM2 to TM4: transmembrane regions 2 to 4.'C' : conserved cysteine. '*' : position of the pattern.

A conserved region that includes two cysteines and seems to be located in a short cytoplasmic loop between two transmembrane domains has been selected as a signature for these proteins.

Consensus pattern: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]- x(2)-[EG]-x(2)-[CWN]-[LIVM](2)

[1] Levy S., Nguyen V.Q., Andria M.L., Takahashi S. J. Biol. Chem. 266:14597-14602(1991).[2] Tomlinson M.G., Williams A.F., Wright M.D. Eur. J. Immunol. 23:136-40(1993).[3] Barclay A.N., Birkeland M.L., Brown M.H., Beyers A.D., Davis S.J., Somoza C., Williams A.F. The leucocyte antigen factbooks. Academic Press, London / San Diego, (1993).

673. Tryptophan synthase alpha chain signature

Tryptophan synthase catalyzes the last step in the biosynthesis of tryptophan: the conversion of indoleglycerol phosphate and serine, totryptophan and glyceraldehyde 3-phosphate [1,2]. It has two functional domains: one for the aldol cleavage of indoleglycerol phosphate to indole andglyceraldehyde 3-phosphate and the other for the synthesis of tryptophan fromindole and serine. In bacteria and plants [3], each domain is found on a separate subunit (alpha and beta chains), while in fungi the two domains are fused together on a single multifunctional protein. A conserved region that contains three conserved acidic residues has been selected as a signature pattern for the alpha chain. The first and the third acidic residues are believed to serve as proton donors/acceptors in the enzyme's catalytic mechanism.

Consensus pattern: [LIVM]-E-[LIVM]-G-x(2)-[FYC]-[ST]-[DE]-[PA]-[LIVMY]- [AGLI]-[DE]-G

[1] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).[2] Hyde C.C., Miles E.W. Bio/Technology 8:27-32(1990).[3] Berlyn M.B., Last R.L., Fink G.R. Proc. Natl. Acad. Sci. U.S.A. 86:4604-4608(1989).

674. Tryptophan synthase beta chain pyridoxal-phosphate attachment site

Tryptophan synthase catalyzes the last step in the biosynthesis of tryptophan: the conversion of indoleglycerol phosphate and serine, totryptophan and glyceraldehyde 3-phosphate [1,2]. It has two functional domains: one for the aldol cleavage of indoleglycerol phosphate to indole andglyceraldehyde 3-phosphate and the other for the synthesis of tryptophan fromindole and serine. In bacteria and plants [3], each domain is found on a separate subunit (alpha and beta chains), while in fungi the two domains arefused together on a single multifunctional protein. The beta chain of the enzyme requires pyridoxal-phosphate as a cofactor. The pyridoxal-phosphate group is attached to a lysine residue. The region around this lysine residue also contains two histidine residues which are part of the pyridoxal-phosphate binding site. The signature pattern for the tryptophansynthase beta chain is derived from that conserved region. -Consensus pattern: [LIVM]-x-H-x-G-[STA]-H-K-x-N [K is the pyridoxal-P attachment site] [1] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).[2] Hyde C.C., Miles E.W. Bio/Technology 8:27-32(1990).[3] Berlyn M.B., Last R.L., Fink G.R. Proc. Natl. Acad. Sci. U.S.A. 86:4604-4608(1989).

675. Serine proteases, trypsin family, active sites

The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases [1]. A partial list of proteases known to belong to the trypsin family is shown below. - Acrosin. - Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C. - Cathepsin G. - Chymotrypsins. - Complement components C1r, C1s, C2, and complement factors B, D and I. - Complement-activating component of RA-reactive factor. - Cytotoxic cell proteases (granzymes A to H). - Duodenase I. - Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin). - Enterokinase (EC 3.4.21.9) (enteropeptidase). - Hepatocyte growth factor activator. - Hepsin. - Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin). - Plasma kallikrein. - Mast cell proteases (MCP) 1 (chymase) to 8. - Myeloblastin (proteinase 3) (Wegener's autoantigen). - Plasminogen activators (urokinase-type, and tissue-type). - Trypsins I, II, III, and IV. - Tryptases. - Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator. - Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab. - Apolipoprotein(a). - Blood fluke

cercarial protease. - Drosophila trypsin like proteases: alpha, easter, snake-locus. - Drosophila protease stubble (gene sb). - Major mite fecal allergen Der p III. All the above proteins belong to family S1 in the classification of peptidases[2,E1] and originate from eukaryotic species. It should be noted thatbacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases are listed below. - Achromobacter lyticus protease I. - Lysobacter alpha-lytic protease. - Streptogrisin A and B (Streptomyces proteases A and B). - Streptomyces griseus glutamyl endopeptidase II. - Streptomyces fradiae proteases 1 and 2.
Consensus pattern: [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue]
Consensus pattern: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue]

[1] Brenner S. Nature 334:528-530(1988).[2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

676. (tsp) Thrombospondin type 1 domain

[1] Bork P; FEBS lett 1993;327:125-130.

677. Tubulin subunits alpha, beta, and gamma signature
Tubulins [1,2], the major constituent of microtubules are dimeric proteins which consist of two closely related subunits (alpha and beta). Tubulin binds two molecules of GTP at two different sites (N and E). At the E (Exchangeable) site, GTP is hydrolyzed during incorporation into the microtubule. Near the E site is an invariant region rich in glycines which is found in both chains andwhich is now [3] said to control the access of the nucleotide to its binding site. A signature pattern was developed from this region. With the exception of the simple eukaryotes, most species express a variety of closely related alpha and beta isotypes. In most species there is a third member of the tubulin family: gamma tubulin. Gamma tubulin is found at microtubule organizing centers (MTOC) such as the spindle poles or the centrosome, suggesting that it is involved in the minus-end nucleation of microtubule assembly [4].
Consensus pattern: [SAG]-G-G-T-G-[SA]-G

[1] Cleveland D.W., Sullivan K.F. Annu. Rev. Biochem. 54:331-365(1985).[2] Joshi H.C., Cleveland D.W. Cell Motil. Cytoskeleton 16:159-163(1990).[3] Hesse J., Thierauf M., Ponstingl H. J. Biol. Chem. 262:15472-15475(1987).[4] Joshi H.C. BioEssays 15:637-643(1993).

Tubulin-beta mRNA autoregulation signal

The stability of beta-tubulin mRNAs are autoregulated by their own translation product [1]. Unpolymerized tubulin subunits bind directly (or activate a factor(s) which binds co-translationally) to the nascent N-terminus of beta-tubulin. This binding is transduced through the adjacent ribosomes to activatean RNAse that degrades the polysome-bound mRNA. The recognition element has been shown to be the first four amino acids of beta-tubulin: Met-Arg-Glu-Ile. Mutations to this sequence abolish the autoregulation effect (except for the replacement of Glu by Asp); transposition of this sequence to an internal region of a polypeptide also suppresses the autoregulatory effect.

Consensus pattern: <M-R-[DE]-[IL]

[1] Cleveland D.W. Trends Biochem. Sci. 13:339-343(1988).

678. (tRNA-synt 2c) Aminoacyl-transfer RNA synthetases class-II signatures. Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]-

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]-

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Delarue M., Moras D. BioEssays 15:675-687(1993).[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991). [5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).[6] Cusack S. Biochimie 75:1077-1081(1993).[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

679. UBA-domain

The UBA-domain (ubiquitin associated domain) is a novel sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway. The structure of the UBA domain consists of a compact three helix bundle [1]. Number of members: 84

[1] Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. Dieckmann T, Withers-Ward ES, Jarosinski MA, Liu CF, Chen IS, Feigon J; Nat Struct Biol 1998;5:1042-1047.

680. UBX domain

Domain present in ubiquitin-regulatory proteins. Present in FAF1 and Shp1p.Number of members: 19

[1] The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. Hofmann K, Bucher P; Trends Biochem Sci 1996;21:172-173.

681. (UCH) Ubiquitin carboxyl-terminal hydrolases family 1 cysteine active site

Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins. There are two distinct families of UCH. The first class consist

of enzymes ofabout 25 Kd and is currently represented by: - Mammalian isozymes L1 and L3. - Yeast YUH1. - Drosophila Uch.One of the active site residues of class-I UCH [3] is a cysteine. A signature pattern has been derived from the region around that residue.

Consensus pattern: Q-x(3)-N-[SA]-C-G-x(3)-[LIVM](2)-H-[SA]-[LIVM]-[SA] [C is the active site residue

[1] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[2] D'andrea A., Pellman D. Crit. Rev. Biochem. Mol. Biol. 33:337-352(1998).[3] Johnston S.C., Larsen C.N., Cook W.J., Wilkinson K.D., Hill C.P. EMBO J. 16:3787-3796(1997).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

682. Ubiquitin carboxyl-terminal hydrolases family 2 signatures (UCH-1)

Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins. There are two distinct families of UCH. The second class consist of largeproteins (800 to 2000 residues) and is currently represented by: - Yeast UBP1, UBP2, UBP3, UBP4 (or DOA4/SSV7), UBP5, UBP7, UBP9, UBP10, UBP11, UBP12, UBP13, UBP14, UBP15 and UBP16. - Human tre-2. - Human isopeptidase T. - Human isopeptidase T-3. - Mammalian Ode-1. - Mammalian Unp. - Mouse Dub-1. - Drosophila fat facets protein (gene faf). - Mammalian faf homolog. - Drosophila D-Ubp-64E. - Caenorhabditis elegans hypothetical protein R10E11.3. - Caenorhabditis elegans hypothetical protein K02C4.3.These proteins only share two regions of similarity. The first region containsa conserved cysteine which is probably implicated in the catalytic mechanism. The second region contains two conserved histidines residues, one of which is also probably implicated in the catalytic mechanism. Signature patterns for both conserved regions have been developed.

Consensus pattern: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-Q [C is the putative active site residue]

Consensus pattern: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y [The two H's are putative active site residues]

[1] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[2] D'andrea A., Pellman D. Crit. Rev. Biochem. Mol. Biol. 33:337-352(1998).[3] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

683. Ubiquitin carboxyl-terminal hydrolases family 2 signatures (UCH-2)

Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins. There are two distinct families of UCH. The second class consist of largeproteins (800 to 2000 residues) and is currently represented by: - Yeast UBP1, UBP2, UBP3, UBP4 (or DOA4/SSV7), UBP5, UBP7, UBP9, UBP10, UBP11, UBP12, UBP13, UBP14, UBP15 and UBP16. - Human tre-2. - Human isopeptidase T. - Human isopeptidase T-3. - Mammalian Ode-1. - Mammalian Unp. - Mouse Dub-1. - Drosophila fat facets protein (gene faf). - Mammalian faf homolog. - Drosophila D-Ubp-64E. - Caenorhabditis elegans hypothetical protein R10E11.3. - Caenorhabditis elegans hypothetical protein K02C4.3.These proteins only share two regions of similarity. The first region containsa conserved cysteine which is probably implicated in the catalytic mechanism. The second region contains two conserved histidines residues, one of which is also probably implicated in the catalytic mechanism. Signature patterns for both conserved regions have been developed.

Consensus pattern: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-Q [C is the putative active site residue]

Consensus pattern: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y [The two H's are putative active site residues]

[1] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[2] D'andrea A., Pellman D. Crit. Rev. Biochem. Mol. Biol. 33:337-352(1998).[3] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

684. UDP-glycosyltransferases signature

UDP glycosyltransferases (UGT) are a superfamily of enzymes that catalyzes the addition of the glycosyl group from a UTP-sugar to a small hydrophobic molecule. This family currently consist of: - Mammalian UDP-glucoronosyl transferases (UDPGT) [1,2]. A large family of membrane-bound microsomal enzymes which catalyze the transfer of glucuronic acid to a wide variety of exogenous and endogenous lipophilic substrates. These enzymes are of major importance in the detoxification and subsequent elimination of xenobiotics such as drugs and

carcinogens. - A large number of putative UDPGT from Caenorhabditis elegans. - Mammalian 2-hydroxyacylsphingosine 1-beta-galactosyltransferase [3] (also known as UDP-galactose-ceramide galactosyltransferase). This enzyme catalyzes the transfer of galactose to ceramide, a key enzymatic step in the biosynthesis of galactocerebrosides, which are abundant sphingolipids of the myelin membrane of the central nervous system and peripheral nervous system. - Plants flavonol O(3)-glucosyltransferase. An enzyme [4] that catalyzes the transfer of glucose from UDP-glucose to a flavanol. This reaction is essential and one of the last steps in anthocyanin pigment biosynthesis. - Baculoviruses ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-) [5] (egt). This enzyme catalyzes the transfer of glucose from UDP-glucose to ectysteroids which are insect molting hormones. The expression of egt in the insect host interferes with the normal insect development by blocking the molting process. - Prokaryotic zeaxanthin glucosyl transferase (gene crtX), an enzyme involved in carotenoid biosynthesis and that catalyses the glycosylation reaction which converts zeaxanthin to zeaxanthin-beta- diglucoside. - Streptomyces macrolide glycosyltransferases [6]. These enzymes specifically inactivates macrolide anitibiotics via 2'-O-glycosylation using UDP-glucose.These enzymes share a conserved domain of about 50 amino acid residues locatedin their C-terminal section and from which a pattern has been extracted todetect them. Consensus pattern: [FW]-x(2)-Q-x(2)-[LIVMYA]-[LIMV]-x(4,6)-[LVGAC]-[LVFYA]-[LIVMF]-[STAGCM]-[HNQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]- [LIVMFA]-x(4)-[PQR]-[LIVMT]-x(3)-[PA]-x(3)-[DES]-[QEHN]

[1] Dutton G.J. (In) Glucoronidation of drugs and other compounds, Dutton G.J., Ed., pp 1-78, CRC Press, Boca Raton, (1980).[2] Burchell B., Nebert D.W., Nelson D.R., Bock K.W., Iyanagi T., Jansen P.L., Lancet D., Mulder G.J., Chowdhury J.R., Siest G., Tephly T.R., Mackenzie P.I. DNA Cell Biol. 10:487-494(1991).[3] Schulte S., Stoffel W. Proc. Natl. Acad. Sci. U.S.A. 90:10265-10269(1993).[4] Furtek D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr. Plant Mol. Biol. 11:473-481(1988).[5] O'Reilly D.R., Miller L.K. Science 245:1110-1112(1989).[6] Hernandez C., Olano C., Mendez C., Salas J.A. Gene 134:139-140(1993).

685. UDP-glucose/GDP-mannose dehydrogenase family

The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catlyze the NAD-dependent 2-fold oxidation of an alcholol to an acid without the release of an aldehyde intermediate [2]. Number of members: 55

[1] Purification and characterization of guanosine diphospho-D-mannose dehydrogenase. A key enzyme in the biosynthesis of alginate by Pseudomonas aeruginosa. Roychoudhury S, May TB, Gill JF, Singh SK, Feingold DS, Chakrabarty AM; J Biol Chem 1989;264:9380-9385. [2] Properties and kinetic analysis of UDP-glucose dehydrogenase from group A streptococci. Irreversible inhibition by UDP-chloroacetol. Campbell RE, Sala RF, van de Rijn I, Tanner ME; J Biol Chem 1997;272:3416-3422.

686. Uracil-DNA glycosylase signature

Uracil-DNA glycosylase (EC 3.2.2.-) (UNG) [1] is a DNA repair enzyme that excises uracil residues from DNA by cleaving the N-glycosylic bond. Uracil in DNA can arise as a result of misincorporation of dUMP residues by DNA polymerase or deamination of cytosine. The sequence of uracil-DNA glycosylase is extremely well conserved [2] in bacteria and eukaryotes as well as in herpes viruses. More distantly related uracil-DNA glycosylases are also found in poxviruses [3].In eukaryotic cells, UNG activity is found in both the nucleus and the mitochondria. Human UNG1 protein is transported to both the mitochondria and the nucleus [4]. The N-terminal 77 amino acids of UNG1 seem to be required for mitochondrial localization [4], but the presence of a mitochondrial transitpeptide has not been directly demonstrated. As a signature for this type of enzyme, the most N-termina conserved region has been selected. This region contains an aspartic acid residue which has been proposed, based on X-ray structures [5,6] to act as a general base in the catalytic mechanism.

Consensus pattern: [KR]-[LIV]-[LIVC]-[LIVM]-x-G-[QI]-D-P-Y [D is the active site residue]-

[1] Sancar A., Sancar G.B. Annu. Rev. Biochem. 57:29-67(1988).[2] Olsen L.C., Aasland R., Wittwer C.U., Krokan H.E., Helland D.E. EMBO J. 8:3121-3125 (1989).[3] Upton C., Stuart D.T., McFadden G. Proc. Natl. Acad. Sci. U.S.A. 90:4518-4522(1993).[4] Slupphaug G., Markussen F.-H., Olsen L.C., Aasland R., Aarsaether N., Bakke O., Krokan H.E., Helland D.E. Nucleic Acids Res. 21:2579-2584(1993).[5] Savva R., McAuley-Hecht K., Brown T., Pearl L. Nature 373:487-493(1995).[6] Mol C.D., Arvai A.S., Slupphaug G., Kavli B., Alseth I., Krohan H.E., Tainer J.A. Cell 80:869-878(1995).[7] Muller S.J., Caradonna S. Biochim. Biophys. Acta 1088:197-207(1991).[8] Meyer-Siegler K., Mauro D.J., Seal G., Wurzer J., Deriel J.K., Sirover M.A. Proc. Natl. Acad. Sci. U.S.A. 88:8460-8464(1991).[9] Muller S.J., Caradonna S. J. Biol. Chem. 268:1310-1319(1993).[10] Barnes D.E., Lindahl T., Sedgwick B. Curr. Opin. Cell Biol. 5:424-433(1993).

687. Uncharacterized protein family UPF0001 signature

The following uncharacterized proteins have been shown [1] to share regions ofsimilarities: - Yeast chromosome II hypothetical protein YBL036c. - Caenorhabditis elegans hypothetical protein F09E5.8. - Bacillus subtilis hypothetical protein ylmE. - Escherichia coli hypothetical protein yggS and HI0090, the corresponding Haemophilus influenzae protein. - Helicobacter pylori hypothetical protein HP0395. - Mycobacterium tuberculosis hypothetical protein MtCY270.20. - Synechocystis strain PCC 6803 hypothetical protein slr0556. - A Pseudomonas aeruginosa hypothetical protein in pilT 5'region. - A Vibrio alginolyticus hypothetical protein in pilT 5'region. These are proteins of from 25 to 30 Kd which contain a number of conserved regions. The best conserved region which is located in the first third of these proteins has been selected as a signature pattern.

Consensus pattern: [FW]-H-[FM]-[IV]-G-x-[LIV]-Q-x-[NKR]-K-x(3)-[LIV]

[1] Bairoch A., Rudd K.E. Unpublished observations (1996).

688. Uncharacterized protein family UPF0003 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli protein aefA. - Escherichia coli hypothetical protein yggB. - Escherichia coli hypothetical protein yjeP and HI0195.1, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein ynaI. - Bacillus subtilis hypothetical protein yhdY. - Helicobacter pylori hypothetical protein HP0415. - Synechocystis strain PCC 6803 hypothetical protein slr0639. - Archaeoglobus fulgidus hypothetical protein AF1546. - Methanococcus jannaschii hypothetical protein MJ0170. - Methanococcus jannaschii hypothetical protein MJ1143.The size of these proteins range from 30 to 120 Kd. They all contain a number of transmembrane regions. The best conserved region which is located in and just after the last potential transmembrane region has been selected as a signature pattern,.

Consensus pattern: G-[STIF]-V-x(2)-[LIVM]-x(6)-[LIVMF]-x(3)-[DQ]-x(3)-[LIV]- x-[LIV]-P-N-x(2)-[LIVMF]-[LIVFSTA]-x(5)-N

[1] Bairoch A. Unpublished observations (1997).

689. Uncharacterized protein family UPF0004 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yliG. - Escherichia coli hypothetical protein yleA and HI0019, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yqeV. - Helicobacter pylori hypothetical protein HP0269. - Helicobacter pylori hypothetical protein HP0285. - Mycoplasma iowae hypothetical protein in 16S RNA 5'region. - Mycobacterium leprae hypothetical protein B2235_C2_195. - Pseudomonas aeruginosa hypothetical protein in hemL 3'region. - Synechocystis strain PCC 6803 hypothetical protein slr0082. - Synechocystis strain PCC 6803 hypothetical protein sll0996. - Methanococcus jannaschii hypothetical protein MJ0865. - Methanococcus jannaschii hypothetical protein MJ0867. - Caenorhabditis elegans hypothetical protein F25B5.5.The size of these proteins range from 47 to 61 Kd. They contain six conserved cysteines, three of which are clustered in a region that can be used as asignature pattern.

Consensus pattern: [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVM]- x(4)-G

[1] Bairoch A. Unpublished observations (1997).

690. Uncharacterized protein family UPF0005 signature

The following proteins seems to be evolutionary related [1]: - Mammalian protein TEGT (Testis Enhanced Gene Transcript). - Escherichia coli hypothetical protein yccA and HI0044, the corresponding Haemophilus influenzae protein. - A probable Pseudomonas aeruginosa ortholog of yccA. These are proteins of about 25 Kd which seem to contain seven transmembranedomains. A signature pattern that corresponds to a region that starts with the beginning of the third transmembrane domain and ends in the middle of the fourth one has been developed.

Consensus pattern: G-[LIVM](2)-[SA]-x(5,8)-G-x(2)-[LIVM]-G-P-x-L-x(4)-[SAG]- x(4,6)-[LIVM](2)-x(2)-A-x(3)-T-A-[LIVM](2)-F

[1] Walter L., Marynen P., Szpirer J., Levan G., Guenther E. Genomics 28:301-304(1995).

691. Uncharacterized protein family UPF0006 signatures

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome II hypothetical protein YBL055c. - Escherichia coli hypothetical protein

ycfH and HI0454, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yigW. - Escherichia coli hypothetical protein yjjV and HI0081, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yabD. - Haemophilus influenzae hypothetical protein HI1664. - Mycoplasma genitalium hypothetical protein MG009. These are proteins of from 24 to 47 Kd which contain a number of conserved regions. They can be picked up in the database by the following patterns.

Consensus pattern: [LIVMFY](2)-D-[STA]-H-x-H-[LIVMF]-[DN

Consensus pattern: P-[LIVM]-x-[LIVM]-H-x-R-x-[TA]-x-[DE

Consensus pattern: [LVSA]-[LIVA]-x(2)-[LIVM]-[PS]-x(3)-L-[LIVM]-[LIVMS]-E-T- D-x-P

[1] Bairoch A., Rudd K.E. Unpublished observations (1995).

692. Uncharacterized protein family UPF0007 signature

The following proteins seems to be evolutionary related [1]: - Escherichia coli hypothetical protein ygbP and HI0672, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yacM. - Mycobacterium tuberculosis hypothetical protein MtCY06G11.29c. - Synechocystis strain PCC 6803 hypothetical protein slr0951. - A Rhodobacter capsulatus hypothetical protein in nifR3 5'region. Except for the Rhodobacter protein which contains a C-terminal extension, all these proteins have from 225 to 236 amino acids. They are hydrophilic proteins that can be picked up in the database by the following pattern.

Consensus pattern: V-L-[IV]-H-D-[GA]-A-R

[1] Bairoch A. Unpublished observations (1997).

693. Uncharacterized protein family UPF0015 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome II hypothetical protein YBR002c. - Yeast chromosome XIII hypothetical protein YMR101c. - Escherichia coli hypothetical protein yaeU and HI0920, the corresponding Haemophilus influenzae protein. - Helicobacter pylori hypothetical protein HP1221. - Mycobacterium leprae hypothetical protein B1937_F2_65. - A Corynebacterium glutamicum hypothetical protein in aroF 3'region. - A Streptomyces fradiae hypothetical protein in transposon Tn4556. - Synechocystis strain PCC 6803 hypothetical protein sll0505.

- Methanococcus jannaschii hypothetical protein MJ1372.These are proteins of about 26 to 40 Kd whose central region is well conserved. They can be picked up in the database by the following pattern.

Consensus pattern: [DE]-[LIVMF](3)-R-T-[SG]-G-x(2)-R-x-S-x-[FY]-[LIVM](2)-W-Q-

[1] Wolfe K.H., Lohan A.J.E. Yeast 10:S41-S46(1994).

694. Uncharacterized protein family UPF0016 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast hypothetical protein YBR187w. - Fission yeast hypothetical protein SpAC17G8.08c. - Mouse protein pFT27. - Synechocystis strain PCC 6803 hypothetical protein sll0615. These are hydrophobic proteins of 200 to 320 amino acids that seem to contain six or seven transmembrane domains. A conserved region which seems, in the eukaryotic proteins of this family, to directly follow the second transmembrane domain has been selected as a signature pattern.

Consensus pattern: E-[LIVM]-G-D-K-T-F-[LIVMF](2)-A-

[1] Bairoch A. Unpublished observations (1996).

695. Uncharacterized protein family UPF0021 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome VII hypothetical protein YGL211w. - Dictyostelium discoideum protein veg136. - Methanococcus jannaschii hypothetical proteins MJ1157 and MJ1478.These are proteins of from 300 to 36o residues. They can be picked up in thedatabase by the following pattern which is located in their N-terminalsection.

Consensus pattern: C-K-x(2)-F-x(4)-E-x(22,23)-S-G-G-K-D

[1] Bairoch A. Unpublished observations (1997).

696. Uncharacterized protein family UPF0023 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Mouse protein 22A3. - Yeast chromosome XII hypothetical protein YLR022c. - Caenorhabditis elegans hypothetical protein W06E11.4. - Methanococcus jannaschii

hypothetical protein MJ0592.These are hydrophilic proteins of about 30 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: D-x-D-E-[LIV]-L-x(4)-V-F-x(3)-S-K-G-

[1] Bairoch A. Unpublished observations (1997).

697. Uncharacterized protein family UPF0024 signature. The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein ygbO and HI0701, the corresponding Haemophilus influenzae protein. - Helicobacter pylori hypothetical protein HP0926. - Yeast chromosome XV hypothetical protein YOR243c. - Caenorhabditis elegans hypothetical protein B0024.11. - Methanococcus jannaschii hypothetical proteins MJ0588 and MJ1364.These are hydrophilic proteins of from 39 to 77 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: G-x-K-D-[KR]-x-A-[LV]-T-x-Q-x-[LIVF]-[SGC]-

[1] Bairoch A. Unpublished observations (1997).

698. Uncharacterized protein family UPF0025 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yfcE. - Bacillus subtilis hypothetical protein ysnB. - Mycoplasma genitalium and pneumoniae hypothetical protein MG207. - Methanococcus jannaschii hypothetical proteins MJ0623 and MJ0936. These are hydrophilic proteins of about 20 Kd. They can be picked up in thedatabase by the following pattern.

Consensus pattern: D-V-[LIV]-x(2)-G-H-[ST]-H-x(12)-[LIVMF]-N-P-G

[1] Bairoch A. Unpublished observations (1997).

699. Uncharacterized protein family UPF0029 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome III hypothetical protein YCR59c. - Yeast chromosome IV hypothetical protein YDL177C. - Escherichia coli hypothetical protein yigZ and HI0722, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yvyE.

- A Thermus aquaticus hypothetical protein in pol 5'region. These proteins can be picked up in the database by the following pattern.

Consensus pattern: G-x(2)-[LIVM](2)-x(2)-[LIVM]-x(4)-[LIVM]-x(5)-[LIVM](2)-x- R-[FYW](2)-G-G-x(2)-[LIVM]-G

[1] Koonin E.V., Bork P., Sander C. EMBO J. 13:493-503(1994).

700. Uncharacterized protein family UPF0030 signature

The following uncharacterized proteins have been shown [1] to be highly similar: - Yeast chromosome VI hypothetical protein YFL060c. - Yeast chromosome XIII hypothetical protein YMR095c. - Yeast chromosome XIV hypothetical protein YNL334c. - Bacillus subtilis hypothetical protein yaaE. - Haemophilus influenzae hypothetical protein HI1648. - Methanococcus jannaschii hypothetical protein MJ1661.These are hydrophilic proteins of about 19 to 25 Kd. They can be picked up inthe database by the following pattern.

Consensus pattern: [GA]-L-I-[LIV]-P-G-G-E-S-T-[STA]

[1] Bairoch A. Unpublished observations (1997).

701. Uncharacterized protein family UPF0032 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yigU and HI0188, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein ycbT. - Mycobacterium tuberculosis hypothetical protein MtCY49.33c and U2126A, the corresponding Mycobacterium leprae protein. - Synechocystis strain PCC 6803 hypothetical protein sll0194. - Odontella sinensis and Porphyra purpurea chlroplast hypothetical protein ycf43.These proteins have from 245 to 317 amino acids and seem to contain at least six or seven transmembrane regions. A conserved region located in the central section of these proteins has been developed as a signature pattern,.

Consensus pattern: Y-x(2)-F-[LIVMA](2)-x-L-x(4)-G-x(2)-F-[EQ]-[LIVMF]-P- [LIVM] –

[1] Bairoch A., Rudd K.E. Unpublished observations (1996).

702. Uncharacterized protein family UPF0034 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yhdG and HI0979, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yjbN and HI0634, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yohI and HI0270, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yacF. - Rhodobacter capsulatus protein nifR3 and related proteins in Azospirillum brasilense and Rhizobium leguminosarum. - Synechocystis strain PCC 6803 hypothetical protein slr0644. - Synechocystis strain PCC 6803 hypothetical protein sll0926. - Caenorhabditis elegans hypothetical protein C45G9.2. - Yeast protein SMM1. - Yeast hypothetical protein YLR401c. - Yeast hypothetical protein YLR405w. - Yeast hypothetical protein YML080w. Although it has been proposed [2] that Rhodobacter capsulatus nifR3 is a transcriptional regulatory protein, it is believed that these proteins constitute a family of enzymes whose active site could include a conserved cysteine which has been used as the central part of a signature pattern.

Consensus pattern: [LIVM]-[DNG]-[LIVM]-N-x-G-C-P-x(3)-[LIVMASQ]-x(5)-G-[SAC]

[1] Bairoch A., Rudd K.E. Unpublished observations (1995).[2] Foster-Hartnett D., Cullen P.J., Gabbert K.K., Kranz R.G. Mol. Microbiol. 8:903-914(1993).

703. Uncharacterized protein family UPF0038 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yacE and HI0890, the corresponding Haemophilus influenzae protein. - Mycobacterium tuberculosis hypothetical protein MtCY01B2.23 and O410, the corresponding Mycobacterium leprae protein. - Synechocystis strain PCC 6803 hypothetical protein slr0553. - Other hypothetical proteins from Aeromonas hydrophila, Bacteroides nodosus, Neisseria gonorrhoeae, Pseudomonas putida, Thermus thermophilus and Xanthomonas campestris. - Human hypothetical protein pOV-2. - Yeast hypothetical protein YDR196C. - Caenorhabditis elegans hypothetical protein T05G5.5.These proteins all contain, in their N-terminal extremity, an ATP/GTP-binding motif 'A' (P-loop) (see <PDOC00017>). The size of these proteins range from 200 to 290 residues (with the exception of the Mycobacterial sequences which are are 410 residues long). A conseved region some 50 residues away from the ATP-binding P-loop has been developed as a signature pattern.

Consensus pattern: G-x-[LI]-x-R-x(2)-L-x(4)-F-x(8)-[LIV]-x(5)-P-x-[LIV] –

[1] Rudd K.E., Bairoch A. Unpublished observations (1997).

704. Ubiquitin-conjugating enzymes active site

Ubiquitin-conjugating enzymes (UBC or E2 enzymes) [1,2,3] catalyze the covalent attachment of ubiquitin to target proteins. An activatedubiquitin moiety is transferred from an ubiquitin-activating enzyme (E1) to E2which later ligates ubiquitin directly to substrate proteins with or without the assistance of 'N-end' recognizing proteins (E3). In most species there are many forms of UBC (at least 9 in yeast) which are implicated in diverse cellular functions. A cysteine residue is required for ubiquitin-thiolester formation. There is a single conserved cysteine in UBC's and the region around that residue isconserved in the sequence of known UBC isozymes. That region has been used as a signature pattern.

Consensus pattern: [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x- [LIV] [C is the active site residue]

[1] Jentsch S., Seufert W., Sommer T., Reins H.-A. Trends Biochem. Sci. 15:195-198(1990).[2] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[3] Hershko A. Trends Biochem. Sci. 16:265-268(1991).

705. Uroporphyrinogen decarboxylase signatures

Uroporphyrinogen decarboxylase (URO-D), the fifth enzyme of the heme biosynthetic pathway, catalyzes the sequential decarboxylation of the four acetyl side chains of uroporphyrinogen to yield coproporphyrinogen [1].URO-D deficiency is responsible for the Human genetic diseases familialporphyria cutanea tarda (fPCT) and hepatoerythropoietic porphyria (HEP).The sequence of URO-D has been well conserved throughout evolution. The best conserved region is located in the N-terminal section; it contains a perfectlyconserved hexapeptide. There are two arginine residues in this hexapeptide which could be involved in the binding, via salt bridges, to the carboxylgroups of the propionate side chains of the substrate. This region has been used as a signature pattern. A second signature pattern is based on a another well conserved region which is located in the central section of the protein.

Consensus pattern: P-x-W-x-M-R-Q-A-G-R

Consensus pattern: G-F-[STAGCV]-[STAGC]-x-P-[FYW]-T-[LV]-x(2)-Y-x(2)-[AE]- [GK]

[1] Garey J.R., Labbe-Bois R., Chelstowska A., Rytka J., Harrison L., Kushner J., Labbe P. Eur. J. Biochem. 205:1011-1016(1992).

706. ubiE/COQ5 methyltransferase family signatures

The following methyltransferases have been shown [1] to share regions of similarities: - Escherichia coli ubiE, which is involved in both ubiquinone and menaquinone biosynthesis and which catalyzes the S-adenosylmethionine dependent methylation of 2-polyprenyl-6-methoxy-1,4-benzoquinol into 2-polyprenyl-3- methyl-6-methoxy-1,4-benzoquinol and of demethylmenaquinol into menaquinol. - Yeast COQ5, a ubiquinone biosynthesis methlytransferase. - Bacillus subtilis spore germination protein C2 (gene: gercB or gerC2), a probable menaquinone biosynthesis methlytransferase. - Lactococcus lactis gerC2 homolog. - Caenorhabditis elegans hypothetical protein ZK652.9. - Leishmania donovani amastigote-specific protein A41.These are hydrophilic proteins of about 30 Kd (except for ZK652.9 which is 65Kd). They can be picked up in the database by the following patterns.

Consensus pattern: Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W

Consensus pattern: R-V-[LIVM]-K-[PV]-G-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S

[1] Lee P.T., Hsu A.Y., Ha H.T., Clarke C.F. J. Bacteriol. 179:1748-1754(1997).

707. Uricase signature

Uricase (urate oxidase) [1] is the peroxisomal enzyme responsible for the degradation of urate into allantoin. Some species, like primates and birds, have lost the gene for uricase and are therefore unable to degradeurate. Uricase is a protein of 300 to 400 amino acids. A highly conserved region located in the central part of the sequence has been used as a signature pattern.

Consensus pattern: [LV]-x-[LV]-[LIV]-K-[STV]-[ST]-x-[SN]-x-F-x(2)-[FY]-x(4)- [FY]-x(2)-L-x(5)-R

[1] Motojima K., Kanaya S., Goto S. J. Biol. Chem. 263:16677-16681(1988).

708. Universal stress protein family (Usp)

By a wide range of stress conditions members of the Usp family are predicted to be related to the MADS-box proteins transcript_fact and bind to DNA [2]. Number of members: 39

[1] Expression and role of the universal stress protein, UspA, of Escherichia coli during growth arrest. Nystrom T, Neidhardt FC; Mol Microbiol 1994; 11:537-544.
[2] Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. Mushegian AR, Koonin EV; Genetics 1996; 144:817-828.

709. Ubiquitin domain signature and profile

Ubiquitin [1,2,3] is a protein of seventy six amino acid residues, found in all eukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins,maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis. In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable (up to twelve in a Xenopus gene). In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat. The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins. Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg- Gly-Gly) extending from the compact structure to form a 'tail', important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins. There are a number of proteins which are evolutionary related to ubiquitin: - Ubiquitin-like proteins from baculoviruses as well as in some strains of bovine viral diarrhea viruses (BVDV). These proteins are highly similar to their eukaryotic counterparts. - Mammalian protein GDX [4]. GDX is composed of two domains, a N-terminal ubiquitin-like domain of 74 residues and a C-terminal domain of 83 residues with some similarity with the thyroglobulin hormonogenic site. - Mammalian protein FAU [5]. FAU is a fusion protein which consist of a N-terminal ubiquitin-like protein of 74 residues fused to ribosomal protein S30. - Mouse protein NEDD-8 [6], a ubiquitin-like protein of 81 residues. - Human protein

BAT3, a large fusion protein of 1132 residues that contains a N-terminal ubiquitin-like domain. - Caenorhabditis elegans protein ubl-1 [7]. Ubl-1 is a fusion protein which consist of a N-terminal ubiquitin-like protein of 70 residues fused to ribosomal protein S27A. - Yeast DNA repair protein RAD23 [8]. RAD23 contains a N-terminal domain that seems to be distantly, yet significantly, related to ubiquitin. - Mammalian RAD23-related proteins RAD23A and RAD23B. - Mammalian BCL-2 binding athanogene-1 (BAG-1). BAG-1 is a protein of 274 residues that contains a central ubiquitin-like domain. - Human spliceosome associated protein 114 (SAP 114 or SF3A120). - Yeast protein DSK2, a protein involved in spindle pole body duplication and which contains a N-terminal ubiquitin-like domain. - Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain a N-terminal ubiquitin domain and a C-terminal CAP-Gly domain. - Schizosaccharomyces pombe hypothetical protein SpAC26A3.16. This protein contains a N-terminal ubiquitin domain. - Yeast protein SMT3. - Human ubiquitin-like proteins SMT3A and SMT3B. - Human ubiquitin-like protein SMT3C (also known as PIC1; Ubl1, Sumo-1; Gmp-1 or Sentrin). This protein is involved in targeting ranGAP1 to the nuclear pore complex protein ranBP2. - SMT3-like proteins in plants and Caenorhabditis elegans. To identify ubiquitin and related proteins, a pattern has been developed based on conserved positions in the central section of the sequence. A profile was also developed that spans the complete length of the ubiquitin domain.

Consensus pattern: K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-[LIVMFY]-x-G-x(4)-[DE]

[1] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[2] Monia B.P., Ecker D.J., Croke S.T. Bio/Technology 8:209-215(1990).[3] Finley D., Varshavsky A. Trends Biochem. Sci. 10:343-347(1985).[4] Filippi M., Tribioli C., Toniolo D. Genomics 7:453-457(1990).[5] Olvera J., Wool I.G. J. Biol. Chem. 268:17967-17974(1993).[6] Kumar S., Yoshida Y., Noda M. Biochem. Biophys. Res. Commun. 195:393-399(1993).[7] Jones D., Candido E.P. J. Biol. Chem. 268:19545-19551(1993).[8] Melnick L., Sherman F. J. Mol. Biol. 233:372-388(1993).

710. VHS domain

Domain present in VPS-27, Hrs and STAM. Number of members: 27

711. Vinculin family signatures

Vinculin [1] is a eukaryotic protein that seems to be involved in the attachment of the actin-based microfilaments to the plasma membrane. Vinculinis located at the cytoplasmic side of focal contacts or adhesion plaques. In addition to actin, vinculin interacts with other structural proteins such as talin and alpha-actinins. Vinculin is a large protein of 116 Kd (about a 1000 residues). Structurally the protein consists of an acidic N-terminal domain of about 90 Kd separated from a basic C-terminal domain of about 25 Kd by a proline-rich region of about 50 residues. The central part of the N-terminal domain consists of avariable number (3 in vertebrates, 2 in Caenorhabditis elegans) of repeats of a 110 amino acids domain. Catenins [2] are proteins that associate with the cytoplasmic domain of avariety of cadherins. The association of catenins to cadherins produces a complex which is linked to the actin filament network, and which seems to be of primary importance for cadherins cell-adhesion properties. Three different types of catenins seem to exist: alpha, beta, and gamma. Alpha-catenins are proteins of about 100 Kd which are evolutionary related to vinculin. Interm of their structure the most significant differences are the absence, inalpha-catenin, of the repeated domain and of the proline-rich segment. Two signature patterns for this family of proteins have been devolped. The first pattern is located in the N-terminal section of both vinculin and alpha-catenins and is part, in vinculin, of a domain that seems to be involved with the interaction with talin. The second pattern is based on a conserved regionin the N-terminal part of the repeated domain of vinculin.

Consensus pattern: [KR]-x-[LIVMF]-x(3)-[LIVMA]-x(2)-[LIVM]-x(6)-R-Q-Q-E-L

Consensus pattern: [LIVM]-x-[QA]-A-x(2)-W-[IL]-x-[DN]-P

[1] Otto J.J. Cell Motil. Cytoskeleton 16:1-6(1990).[2] Herrenknecht K., Ozawa M., Eckerskorn C., Lottspeich F., Lenter M., Kemler R. Proc. Natl. Acad. Sci. U.S.A. 88:9156-9160(1991).

712. (Vitellogenin N) Lipoprotein amino terminal region

This family contains regions from: Vitellogenin, Microsomal triglyceride transfer protein and apolipoprotein B-100. These proteins are all involved in lipid transport [1]. This family contains the LV1n chain from lipovitellin, that contains two structural domains.

Number of members: 33

[1] The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. Anderson TA, Levitt DG, Banaszak LJ Structure 1998;6:895-909.

713. (VMSA) Major surface antigen from hepadnavirus

714. ssDNA binding protein (Viral DNA bp)
This protein is found in herpesviruses and is needed for replication.

715. (Votage CLC) Voltage gated chloride channels

This family of ion channels contains 10 or 12 transmembrane helices. Each protein forms a single pore. It has been shown that some members of this family form homodimers. These proteins contain two CBS domains.

[1] Schmidt-Rose T, Jentsch TJ; J Biol Chem 1997;272:20515-20521.
[2] Zhang J, George AL Jr, Griggs RC, Fouad GT, Roberts J, Kwiecinski H, Connolly AM, Ptacek LJ; Neurology 1996;47:993-998.

716. von Willebrand factor type A domain (vwa)
More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor.
Bork P, Rohde K;
Biochem J 1991;279:908-911.

1. RUGGERI, Z.M. and WARE, J.
von Willebrand factor.
FASEB J. 7 308-316 (1993).

2. COLOMBATTI, A., BONALDO, P. and DOLIANA, R.
Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins.
MATRIX 13 297-306 (1993).

3. PERKINS, S.J., SMITH, K.F., WILLIAMS, S.C., HARIS, P.I., CHAPMAN, D. and SIM, R.B.
The secondary structure of the von Willebrand factor type A domain in factor B of human complement by Fourier transform infrared spectroscopy. Its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged structure predictions.
J.MOL.BIOL. 238 104-119 (1994).

4. BORK, P. and ROHDE, K.
More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor.
BIOCHEM.J. 279 908-910 (1991).

5. EDWARDS, Y.J.K. and PERKINS, S.J.
The protein fold of the von Willebrand factor type A domain is predicted to be similar to the open twisted beta-sheet flanked by alpha-helices found in human ras-p21.
FEBS LETT. 358 283-286 (1995).

6. LEE, J.O., RIEU, P., ARNAOUT, M.A. and LIDDINGTON, R.
Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18).
CELL 80 631-638 (1995).

7. QU, A. and LEAHY, D.J.
Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha L beta 2) integrin.

PROC.NATL.ACAD.SCI.USA 92 10277-10281 (1995).

The von Willebrand factor is a large multimeric glycoprotein found in blood plasma. Mutant forms are involved in the aetiology of bleeding disorders [1]. In von Willebrand factor, the type A domain (vWF) is the prototype for a protein superfamily. The vWF domain is found in various plasma proteins: complement factors B, C2, CR3 and CR4; the integrins (I-domains); collagen types VI, VII, XII and XIV; and other extracellular proteins [2-4]. Proteins that incorporate vWF domains participate in numerous biological events (e.g., cell adhesion, migration, homing, pattern formation, and signal transduction), involving interaction with a large array of ligands [2]. Secondary structure prediction from 75 aligned vWF sequences has revealed a largely alternating sequence of alpha-helices and beta-strands [3]. Fold recognition algorithms were used to score sequence compatibility with a library of known structures: the vWF domain fold was predicted to be a doubly-wound, open, twisted beta-sheet flanked by alpha-helices [5]. 3D structures have been determined for the I-domains of integrins CD11b (with bound magnesium) [6] and CD11a (with bound manganese) [7]. The domain adopts a classic alpha/beta Rossmann fold and contains an unusual metal ion coordination site at its surface. It has been suggested that this site represents a general metal ion-dependent adhesion site (MIDAS) for binding protein ligands [6]. The residues constituting the MIDAS motif in the CD11b and CD11a I-domains are completely conserved, but the manner in which the metal ion is coordinated differs slightly [7].

VWFADOMAIN is a 3-element fingerprint that provides a signature for the vWF domain superfamily. The fingerprint was derived from an initial alignment of 14 sequences. Motif 1 includes the first beta-strand and 3 conserved residues involved in metal ion coordination in I-domains (Asp and 2 serines in positions 8, 10 and 12, respectively); motif 2 spans strands beta-2 and beta-2'; and motif 3 encodes beta-strand 3 and a conserved Asp (in position 7), which coordinates the metal ion [6,7]. Three iterations on OWL27.0 were required to reach convergence, at which point a true set comprising 56 sequences was identified. Numerous partial matches were also found.

717. (WD40) WD domain, G-beta repeat

The ancient regulatory-protein family of WD-repeat proteins.
Neer EJ, Schmidt CJ, Nambudripad R, Smith TF;
Nature 1994;371:297-300.

Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors [1]. The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

In higher eukaryotes G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). Such a repetitive segment has been shown [E1,2,3,4,5] to exist in a number of other proteins listed below:

- Yeast STE4, a component of the pheromone response pathway. STE4 is a G-beta like protein that associates with GPA1 (G-alpha) and STE18 (G-gamma).
- Yeast MSI1, a negative regulator of RAS-mediated cAMP synthesis. MSI1 is most probably also a G-beta protein.

- Human and chicken protein 12.3. The function of this protein is not known, but on the basis of its similarity to G-beta proteins, it may also function in signal transduction.
- Chlamydomonas reinhardtii gblp. This protein is most probably the homolog of vertebrate protein 12.3.
- Human LIS1, a neuronal protein involved in type-1 lissencephaly [E2].
- Mammalian coatomer beta' subunit (beta'-COP), a component of a cytosolic protein complex that reversibly associates with Golgi membranes to form

vesicles that mediate biosynthetic protein transport.

- Yeast CDC4, essential for initiation of DNA replication and separation of the spindle pole bodies to form the poles of the mitotic spindle.
- Yeast CDC20, a protein required for two microtubule-dependent processes: nuclear movements prior to anaphase and chromosome separation.
- Yeast MAK11, essential for cell growth and for the replication of M1 double-stranded RNA.
- Yeast PRP4, a component of the U4/U6 small nuclear ribonucleoprotein with a probable role in mRNA splicing.
- Yeast PWP1, a protein of unknown function.
- Yeast SKI8, a protein essential for controlling the propagation of double-stranded RNA.
- Yeast SOF1, a protein required for ribosomal RNA processing which associates with U3 small nucleolar RNA.
- Yeast TUP1 (also known as AER2 or SFL2 or CYC9), a protein which has been implicated in dTMP uptake, catabolite repression, mating sterility, and many other phenotypes.
- Yeast YCR57c, an ORF of unknown function from chromosome III.
- Yeast YCR72c, an ORF of unknown function from chromosome III.

- Slime mold coronin, an actin-binding protein.
- Slime mold AAC3, a developmentally regulated protein of unknown function.

- Drosophila protein Groucho (formerly known as E(spl); 'enhancer of split'), a protein involved in neurogenesis and that seems to interact with the Notch and Delta proteins.
- Drosophila TAF-II-80, a protein that is tightly associated with TFIID.

The number of repeats in the above proteins varies between 5 (PRP4, TUP1, and Groucho) and 8 (G-beta, STE4, MSI1, AAC3, CDC4, PWP1, etc.). In G-beta and G-beta like proteins, the repeats span the entire length of the sequence, while in other proteins, they make up the N-terminal, the central or the C-terminal section.

A signature pattern can be developed from the central core of the domain (positions 9 to 23).

-Consensus pattern: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN]

[1] Gilman A.G.
Annu. Rev. Biochem. 56:615-649(1987).
[2] Duronio R.J., Gordon J.I., Boguski M.S.
Proteins 13:41-56(1992).
[3] van der Voorn L., Ploegh H.L.
FEBS Lett. 307:131-134(1992).
[4] Neer E.J., Schmidt C.J., Nambudripad R., Smith T.F.
Nature 371:297-300(1994).
[5] Smith T.F., Gaiatzes C.G., Saxena K., Neer E.J.
Biochemistry In Press(1998).

718. WHEP-TRS domain containing proteins

A conserved domain of 46 amino acids has been shown [1] to exist in a number of higher eukaryote aminoacyl-transfer RNA synthetases. This domain is present one to six times in the following enzymes:

- Mammalian multifunctional aminoacyl-tRNA synthetase. The domain is present three times in a region that separates the N-terminal glutamyl-tRNA synthetase domain from the C-terminal prolyl-tRNA synthetase domain.
- Drosophila multifunctional aminoacyl-tRNA synthetase. The domain is present six times in the intercatalytic region.
- Mammalian tryptophanyl-tRNA synthetase. The domain is found at the N-terminal extremity.
- Mammalian, insect, nematode and plant glycyl-tRNA synthetase. The domain is found at the N-terminal extremity [2].

- Mammalian histidyl-tRNA synthetase. The domain is found at the N-terminal extremity.

This domain, which is called WHEP-TRS, could contain a central alpha-helical region and may play a role in the association of tRNA-synthetases into multienzyme complexes.

A signature pattern based on the first 29 positions of the WHEP-Domain has been developed.

-Consensus pattern: [QY]-G-[DNEA]-x-[LIV]-[KR]-x(2)-K-x(2)-[KRNG]-[AS]-x(4)-[LIV]-[DENK]-x(2)-[IV]-x(2)-L-x(3)-K

[1] Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M. EMBO J. 10:4267-4277(1991).
[2] Nada S., Chang P.K., Dignam J.D. J. Biol. Chem. 268:7660-7667(1993).

719. (Worm family 8) Putative membrane protein
Analysis of protein domain families in Caenorhabditis elegans.
Sonnhammer EL, Durbin R;
Genomics 1997;46:200-216.
This family called family 8 in [1], may be a transmembrane protein
The specific function of this protein is unknown.

720. Xylose isomerase
Xylose isomerase (EC 5.3.1.5) [1] is an enzyme found in microorganisms which catalyzes the interconversion of D-xylose to D-xylulose. It can also isomerize D-ribose to D-ribulose and D-glucose to D-fructose. Xylose isomerase seems to require magnesium for its activity, while cobalt is necessary to stabilize the tetrameric structure of the enzyme. A number of residues are conserved in all known xylose isomerases.

Xylose isomerase also exists in plants [2] where it is homodimeric and is manganese-dependent.

Two signatures patterns for xylose isomerase have been developed. The first one is derived from a stretch of five conserved amino acids that includes a glutamic acid residue known to be one of the four residues involved in the binding of the magnesium ion [3]; this pattern also includes a lysine residue which is involved in the catalytic activity. The second pattern is derived from a conserved region in the N-terminal section of the enzyme that include an histidine residue which has been shown [4] to be involved in the catalytic mechanism of the enzyme.

-Consensus pattern: [LI]-E-P-K-P-x(2)-P
[E is a magnesium ligand]
[K is an active site residue]
-Consensus pattern: [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE]
[H is an active site residue]

[1] Dauter Z., Dauter M., Hemker J., Witzel H., Wilson K.S.
FEBS Lett. 247:1-8(1989).
[2] Kristo P.A., Saarelainen R., Fagerstrom R., Aho S., Korhola M.
Eur. J. Biochem. 237:240-246(1996).
[3] Henrick K., Collyer C.A., Blow D.M.
J. Mol. Biol. 208:129-157(1989).
[4] Vangrysperre W., Ampe C., Kersters-Hilderson H., Tempst P.
Biochem. J. 263:195-199(1989).

721. XPG protein signatures. Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can

be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].XPG belongs to a family of proteins [2,3,4,5,6] that are composed of twomain subsets: - Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9]. - Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease. In addition to the proteins listed in the above groups, this family also includes: - Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs. - Yeast EXO1 (DHS1), a protein with probably the same function as exo1. - Yeast DIN7.Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset. Two signature patterns have been developed for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide

Consensus pattern: [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K-
Consensus pattern: [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM]-

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).[2] Scherly D., Nouspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993).[3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993).[4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994).[5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994).[6] Szankasi P., Smith G.R. Science 267:1166-1169(1995).[7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993).[8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem.

269:15965-15968(1994).[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).

722. Xanthine/uracil permeases family

The following transport proteins which are involved in the uptake of xanthine or uracil are evolutionary related [1]:

- Uric uric acid-xanthine permease (gene uapA) from Aspergillus nidulans.
- Purine permease (gene uapC) from Aspergillus nidulans.
- Xanthine permease from Bacillus subtilis (gene pbuX).
- Uracil permease from Escherichia coli (gene uraA) [2] and Bacillus (gene pyrP).
- Hypothetical protein ycdG from Escherichia coli.
- Hypothetical protein ygfO from Escherichia coli.
- Hypothetical protein ygfU from Escherichia coli.
- Hypothetical protein yicE from Escherichia coli.
- Hypothetical protein yunJ from Bacillus subtilis.
- Hypothetical protein yunK from Bacillus subtilis.

They are proteins of from 430 to 595 residues that seem to contain 12 transmembrane domains.

The best conserved region which corresponds with what seems to be the tenth transmembrane domain has been selected as a signature pattern.

-Consensus pattern: [LIVM]-P-x-[PASIF]-V-[LIVM]-G-G-x(4)-[LIVM]-[FY]-[GSA]-x-[LIVM]-x(3)-G

[1] Diallinas G., Gorfinkiel L., Arst G., Cecchetto G., Scazzocchio C. J. Biol. Chem. 270:8610-8622(1995).

[2] Andersen P.S., Frees D., Fast R., Mygind B. J. Bacteriol. 177:2008-2013(1995).

723. Hypothetical yabO/yceC/sfhB family

The following proteins, which seems to belong to a family of pseudouridine synthases (EC 4.2.1.70) [1] have been shown to share regions of similarities:

- Escherichia coli and Haemophilus influenzae ribosomal large subunit pseudouridine synthase A (gene rluA). It is responsible for synthesis of pseudouridine from uracil-746 IN 23S rRNA.
- Escherichia coli and Haemophilus influenzae ribosomal large subunit pseudouridine synthase C (gene rluC). It is responsible for synthesis of pseudouridine from uracil at positions 955, 2504 and 2580 in 23S rRNA.
- Escherichia coli protein and homologs in other bacteria large subunit pseudouridine synthase D (gene rluD).
- Yeast DRAP deaminase (gene RIB2).
- Escherichia coli hypothetical protein yqcB and HI1435, the corresponding Haemophilus influenzae protein.
- Haemophilus influenzae hypothetical protein HI0042.
- Aquifex aeolicus hypothetical protein AQ_1758.
- Bacillus subtilis hypothetical protein yhcT.
- Bacillus subtilis hypothetical protein yjbO.
- Bacillus subtilis hypothetical protein ylyB.
- Helicobacter pylori hypothetical protein HP0347.
- Helicobacter pylori hypothetical protein HP0745.
- Helicobacter pylori hypothetical protein HP0956.
- Mycoplasma genitalium hypothetical protein MG209.
- Mycoplasma genitalium hypothetical protein MG370.
- Synechocystis strain PCC 6803 hypothetical protein slr1592.
- Synechocystis strain PCC 6803 hypothetical protein slr1629.
- Yeast hypothetical protein YDL036c.
- Yeast hypothetical protein YGR169c.
- Fission yeast hypothetical protein SpAC18B11.02c.
- Caenorhabditis elegans hypothetical protein K07E8.7.

These are proteins of from 21 to 50 Kd which contain a number of conserved regions in their central section. They can be picked up in the database by the following highly conserved pattern.

-Consensus pattern: [LIVCA]-[NHYT]-R-[LI]-D-x(2)-T-[STA]-G-[LIVAGC]-[LIVMF](2)-[LIVMFGC]-[SGTACV]

[1] Conrad J., Sun D., Englund N., Ofengand J.
J. Biol. Chem. 273:18562-18566(1998).

In addition, the following bacterial proteins, which seems to belong to a family of pseudouridine synthases (EC 4.2.1.70) [1] also have been shown to share regions of similarities:

- Escherichia coli and Haemophilus influenzae 16S pseudouridylate 516 synthase (EC 4.2.1.70) (gene: rsuA). This enzyme is responsible for the formation of pseudouridine from uracil-516 in 16S ribosomal RNA.
- Escherichia coli hypothetical protein yciL and HI1199, the corresponding Haemophilus influenzae protein.
- Escherichia coli hypothetical protein yjbC.
- Escherichia coli hypothetical protein ymfC and HI0694, the corresponding Haemophilus influenzae protein.
- Aquifex aeolicus hypothetical protein AQ_554.
- Aquifex aeolicus hypothetical protein AQ_1464.
- Bacillus subtilis hypothetical protein ypuL.
- Bacillus subtilis hypothetical protein ytzF.
- Borrelia burgdorferi hypothetical protein BB0129.
- Helicobacter pylori hypothetical protein HP1459.
- Synechocystis strain PCC 6803 hypothetical protein slr0361.
- Synechocystis strain PCC 6803 hypothetical protein slr0612.

These are proteins of from 25 to 40 Kd which contain a number of conserved regions in their central section. They can be picked up in the database by the following highly conserved pattern.

-Consensus pattern: G-R-L-D-x(2)-[STA]-x-G-[LIVFA]-[LIVMF](3)-[ST]-[DNST]

[1] Wrzesinski J., Bakin A., Nurse K., Lane B.G., Ofengand J. Biochemistry 34:8904-8913(1995).

724. Zinc finger present in dystrophin, CBP/p300
ZZ in dystrophin binds calmodulin
Putative zinc finger; binding not yet shown.

725. Zinc carboxypeptidase
There are a number of different types of zinc-dependent carboxypeptidases (EC 3.4.17.-) [1,2]. All these enzymes seem to be structurally and functionally related. The enzymes that belong to this family are listed below.

- Carboxypeptidase A1 (EC 3.4.17.1), a pancreatic digestive enzyme that can removes all C-terminal amino acids with the exception of Arg, Lys and Pro.
- Carboxypeptidase A2 (EC 3.4.17.15), a pancreatic digestive enzyme with a specificity similar to that of carboxypeptidase A1, but with a preference for bulkier C-terminal residues.
- Carboxypeptidase B (EC 3.4.17.2), also a pancreatic digestive enzyme, but that preferentially removes C-terminal Arg and Lys.
- Carboxypeptidase N (EC 3.4.17.3) (also known as arginine carboxypeptidase), a plasma enzyme which protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation.
- Carboxypeptidase H (EC 3.4.17.10) (also known as enkephalin convertase or carboxypeptidase E), an enzyme located in secretory granules of pancreatic islets, adrenal gland, pituitary and brain. This enzyme removes residual C-terminal Arg or Lys remaining after initial endoprotease cleavage during prohormone processing.
- Carboxypeptidase M (EC 3.4.17.12), a membrane bound Arg and Lys specific enzyme.
It is ideally situated to act on peptide hormones at local tissue sites where it could control their activity before or after interaction with

specific plasma membrane receptors.

- Mast cell carboxypeptidase (EC 3.4.17.1), an enzyme with a specificity to carboxypeptidase A, but found in the secretory granules of mast cells.
- Streptomyces griseus carboxypeptidase (Cpase SG) (EC 3.4.17.-) [3], which combines the specificities of mammalian carboxypeptidases A and B.
- Thermoactinomyces vulgaris carboxypeptidase T (EC 3.4.17.18) (CPT) [4], which also combines the specificities of carboxypeptidases A and B.
- AEBP1 [5], a transcriptional repressor active in preadipocytes. AEBP1 seems to regulate transcription by cleavage of other transcriptional proteins.
- Yeast hypothetical protein YHR132c.

All of these enzymes bind an atom of zinc. Three conserved residues are implicated in the binding of the zinc atom: two histidines and a glutamic acid Two signature patterns which contain these three zinc-ligands have been derived.

-Consensus pattern: [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-[LIVMFYTA]
[H and E are zinc ligands]
-Consensus pattern: H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW]
[H is a zinc ligand]

[1] Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A.
J. Biol. Chem. 264:13165-13170(1989).
[2] Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F., Serafin W.E.
J. Biol. Chem. 264:20094-20099(1989).
[3] Narahashi Y.
J. Biochem. 107:879-886(1990).
[4] Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I., Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P., Galperina O.V., Matz M.V., Stepanov V.M.
Eur. J. Biochem. 208:281-288(1992).
[5] He G.-P., Muise A., Li A.W., Ro H.-S.
Nature 378:92-96(1995).

[6] Hourdou M.-L., Guinand M., Vacheron M.J., Michel G., Denoroy L., Duez C.M., Englebert S., Joris B., Weber G., Ghuysen J.-M. Biochem. J. 292:563-570(1993).

[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

726. Zinc finger, C2H2 type

The C2H2 zinc finger is the classical zinc finger domain. The two conserved cysteines and histidines co-ordinate a zinc ion. The following pattern describes the zinc finger.

#-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C]

Where X can be any amino acid, and numbers in brackets indicate the number of residues. The positions marked # are those that are important for the stable fold of the zinc finger. The final position can be either his or cys.

The C2H2 zinc finger is composed of two short beta strands followed by an alpha helix. The amino terminal part of the helix binds the major groove in DNA binding zinc fingers.

'Zinc finger' domains [1-5] are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:

```
  x x
  x   x
 x x
 x x
 x x
```

```
x x
 C   H
x \ /  x
 x  Zn  x
x /   \ x
 C   H
 x x x x x  x x x x x
```

Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

Some of the proteins known to include C2H2-type zinc fingers are listed below. The number of zinc finger regions found in each of these proteins are indicated between brackets; a '+' symbol indicates that only partial sequence data is available and that additional finger domains may be present.

- Saccharomyces cerevisiae: ACE2 (3), ADR1 (2), AZF1 (4), FZF1 (5), MIG1 (2), MSN2 (2), MSN4 (2), RGM1 (2), RIM1 (3), RME1 (3), SFP1 (2), SSL1 (1), STP1 (3), SWI5 (3), VAC1 (1) and ZMS1 (2).
- Emericella nidulans: brlA (2), creA (2).
- Drosophila: AEF-1 (4), Cf2 (7), ci-D (5), Disconnected (2), Escargot (5), Glass (5), Hunchback (6), Kruppel (5), Kruppel-H (4+), Odd-skipped (4), Odd-paired (4), Pep (3), Snail (5), Spalt-major (7), Serependity locus beta (6), delta (7), h-1 (8), Suppressor of hairy wing su(Hw) (12), Suppressor of variegation suvar(3)7 (5), Teashirt (3) and Tramtrack (2).
- Xenopus: transcription factor TFIIIA (9), p43 from RNP particle (9), Xfin (37 !!), Xsna (5), gastrula XlcGF5.1 to XlcGF71.1 (from 4+ to 11+), Oocyte XlcOF2 to XlcOF22 (from 7 to 12).
- Mammalian: basonuclin (6), BCL-6/LAZ-3 (6), erythroid krueppel-like transcription factor (3), transcription factors Sp1 (3), Sp2 (3), Sp3 (3)

and Sp(4) 3, transcriptional repressor YY1 (4), Wilms' tumor protein (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), ZfX (13), ZfY (13), Zfp-35 (18), ZNF7 (15), ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3).

In addition to the conserved zinc ligand residues it has been shown [6] that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue.

-Consensus pattern: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H
[The two C's and two H's are zinc ligands]

[1] Klug A., Rhodes D.
Trends Biochem. Sci. 12:464-469(1987).
[2] Evans R.M., Hollenberg S.M.
Cell 52:1-3(1988).
[3] Payre F., Vincent A.
FEBS Lett. 234:245-250(1988).
[4] Miller J., McLachlan A.D., Klug A.
EMBO J. 4:1609-1614(1985).
[5] Berg J.M.
Proc. Natl. Acad. Sci. U.S.A. 85:99-102(1988).
[6] Rosenfeld R., Margalit H.
J. Biomol. Struct. Dyn. 11:557-570(1993).

727. Zinc finger, C3HC4 type (RING finger)

A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or 'RING' finger) [1] that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D structure of the zinc ligation system is unique

to the RING domain and is refered to as the "cross-brace" motif. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C.

Proteins currently known to include the C3HC4 domain are listed below (references are only provided for recently determined sequences).

- Mammalian V(D)J recombination activating protein (gene RAG1). RAG1 activates the rearrangement of immunoglobulin and T-cell receptor genes.
- Mouse rpt-1. Rpt-1 is a trans-acting factor that regulates gene expression directed by the promoter region of the interleukin-2 receptor alpha chain or the LTR promoter region of HIV-1.
- Human rfp. Rfp is a developmentally regulated protein that may function in male germ cell development. Recombination of the N-terminal section of rfp with a protein tyrosine kinase produces the ret transforming protein.
- Human 52 Kd Ro/SS-A protein. A protein of unknown function from the Ro/SS-A ribonucleoprotein complex. Sera from patients with systemic lupus erythematosus or primary Sjogren's syndrome often contain antibodies that react with the Ro proteins.
- Human histocompatibility locus protein RING1.
- Human PML, a probable transcription factor. Chromosomal translocation of PML with retinoic receptor alpha creates a fusion protein which is the cause of acute promyelocytic leukemia (APL).
- Mammalian breast cancer type 1 susceptibility protein (BRCA1) [E1].
- Mammalian cbl proto-oncogene.
- Mammalian bmi-1 proto-oncogene.
- Vertebrate CDK-activating kinase (CAK) assembly factor MAT1, a protein that stabilizes the complex between the CDK7 kinase and cyclin H (MAT1 stands for 'Menage A Trois').
- Mammalian mel-18 protein. Mel-18 which is expressed in a variety of tumor cells is a transcriptional repressor that recognizes and bind a specific DNA sequence.
- Mammalian peroxisome assembly factor-1 (PAF-1) (PMP35), which is somewhat involved in the biogenesis of peroxisomes. In humans, defects in PAF-1 are

responsible for a form of Zellweger syndrome, an autosomal recessive disorder associated with peroxisomal deficiencies.

- Human MAT1 protein, which interacts with the CDK7-cyclin H complex.
- Human RING1 protein.
- Xenopus XNF7 protein, a probable transcription factor.
- Trypanosoma protein ESAG-8 (T-LR), which may be involved in the postranscriptional regulation of genes in VSG expression sites or may interact with adenylate cyclase to regulate its activity.
- Drosophila proteins Posterior Sex Combs (Psc) and Suppressor two of zeste (Su(z)2). The two proteins belong to the Polycomb group of genes needed to maintain the segment-specific repression of homeotic selector genes.
- Drosophila protein male-specific msl-2, a DNA-binding protein which is involved in X chromosome dosage compensation (the elevation of transcription of the male single X chromosome).
- Arabidopsis thaliana protein COP1 which is involved in the regulation of photomorphogenesis.
- Fungal DNA repair proteins RAD5, RAD16, RAD18 and rad8.
- Herpesviruses trans-acting transcriptional protein ICP0/IE110. This protein which has been characterized in many different herpesviruses is a trans-activator and/or -repressor of the expression of many viral and cellular promoters.
- Baculoviruses protein CG30.
- Baculoviruses major immediate early protein (PE-38).
- Baculoviruses immediate-early regulatory protein IE-N/IE-2.
- Caenorhabditis elegans hypothetical proteins F54G8.4, R05D3.4 and T02C1.1.
- Yeast hypothetical proteins YER116c and YKR017c.

The central region of the domain was selected as a signature pattern for the C3HC4 finger.

-Consensus pattern: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA]

[1] Borden K.L.B., Freemont P.S.
Curr. Opin. Struct. Biol. 6:395-401(1996).

728. Zinc finger C-x8-C-x5-C-x3-H type (and similar).

729. Zinc finger, CCHC class

A family of CCHC zinc fingers, mostly from retroviral gag proteins (nucleocapsid). Prototype structure is from HIV. Also contains members involved in eukaryotic gene regulation, such as C. elegans GLH-1.

Structure is an 18-residue zinc finger; no examples of indels in the alignment.

730. Zn-finger in Ran binding protein and others.

731. AN1-like Zinc finger

Zinc finger at the C-terminus of An1 Swiss:Q91889, a ubiquitin-like protein in Xenopus laevis. The following pattern describes the zinc finger. C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C Where X can be any amino acid, and numbers in brackets indicate the number of residues.

[1] Linnen JM, Bailey CP, Weeks DL; Gene 1993;128:181-188.

732. 14-3-3 proteins

Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways.

Xiao B, Smerdon SJ, Jones DH, Dodson GG, Soneji Y, Aitken A, Gamblin SJ; Nature 1995;376:188-191.

Crystal structure of the zeta isoform of the 14-3-3 protein.

Liu D, Bienkowska J, Petosa C, Collier RJ, Fu H, Liddington R;

Nature 1995;376:191-194.

Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine.
Muslin AJ, Tanner JW, Allen PM, Shaw AS;
Cell 1996;84:889-897.

The 14-3-3 protein binds its target proteins with a common site located towards the C-terminus.
Ichimura T, Ito M, Itagaki C, Takahashi M, Horigome T, Omata S, Ohno S, Isobe T
FEBS Lett 1997;413:273-276.

Molecular evolution of the 14-3-3 protein family.
Wang W, Shakes DC
J Mol Evol 1996;43:384-398.
Function of 14-3-3 proteins.
Jin DY, Lyu MS, Kozak CA, Jeang KT
Nature 1996;382:308-308.

The 14-3-3 proteins [1,2,3] are a family of closely related acidic homodimeric proteins of about 30 Kd which were first identified as being very abundant in mammalian brain tissues and located preferentially in neurons. The 14-3-3 proteins seem to have multiple biological activities and play a key role in signal transduction pathways and the cell cycle. They interacts with kinases such as PKC or Raf-1; they seem to also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases and in plants they are associated with a complex that binds to the G-box promoter elements.

The 14-3-3 family of proteins are ubiquitously found in all eukaryotic species studied and have been sequenced in fungi (yeast BMH1 and BMH2, fission yeast rad24 and rad25), plants, Drosophila, and vertebrates. The sequences of the 14-3-3 proteins are extremely well conserved. Two highly conserved regions have been selected as signature patterns: the first is a peptide of 11 residues

located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section.

-Consensus pattern: R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA]

-Consensus pattern: Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD]

[1] Aitken A.
Trends Biochem. Sci. 20:95-97(1995).

[2] Morrison D.
Science 266:56-57(1994).

[3] Xiao B., Smerdon S.J., Jones D.H., Dodson G.G., Soneji Y., Aitken A., Gamblin S.J.
Nature 376:188-191(1995).

733. D-isomer specific 2-hydroxyacid dehydrogenases (2 Hacid DH)

This Pfam covers the Formate dehydrogenase, D-glycerate dehydrogenase and D-lactate dehydrogenase families in SCOP. A number of NAD-dependent 2-hydroxyacid dehydrogenases which seem to be specific for the D-isomer of their substrate have been shown [1,2,3,4] to be functionally and structurally related. These enzymes are listed below.

- D-lactate dehydrogenase (EC 1.1.1.28), a bacterial enzyme which catalyzes the reduction of D-lactate to pyruvate.
- D-glycerate dehydrogenase (EC 1.1.1.29) (NADH-dependent hydroxypyruvate reductase), a plant leaf peroxisomal enzyme that catalyzes the reduction of hydroxypyruvate to glycerate. This reaction is part of the glycolate pathway of photorespiration.
- D-glycerate dehydrogenase from the bacteria Hyphomicrobium methylovorum and Methylobacterium extorquens.
- 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), a bacterial enzyme that catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. This reaction is the first committed step in the 'phosphorylated' pathway of serine biosynthesis.

- Erythronate-4-phosphate dehydrogenase (EC 1.1.1.-) (gene pdxB), a bacterial enzyme involved in the biosynthesis of pyridoxine (vitamin B6).
- D-2-hydroxyisocaproate dehydrogenase (EC 1.1.1.-) (D-hicDH), a bacterial enzyme that catalyzes the reversible and stereospecific interconversion between 2-ketocarboxylic acids and D-2-hydroxy-carboxylic acids.
- Formate dehydrogenase (EC 1.2.1.2) (FDH) from the bacteria Pseudomonas sp. 101 and various fungi [5].
- Vancomycin resistance protein vanH from Enterococcus faecium; this protein is a D-specific alpha-keto acid dehydrogenase involved in the formation of a peptidoglycan which does not terminate by D-alanine thus preventing vancomycin binding.
- Escherichia coli hypothetical protein ycdW.
- Escherichia coli hypothetical protein yiaE.
- Haemophilus influenzae hypothetical protein HI1556.
- Yeast hypothetical protein YER081w.
- Yeast hypothetical protein YIL074w.

All these enzymes have similar enzymatic activities and are structurally related. Three of the most conserved regions of these proteins have been selected to develop patterns. The first pattern is based on a glycine-rich region located in the central section of these enzymes; this region probably corresponds to the NAD-binding domain. The two other patterns contain a number of conserved charged residues, some of which may play a role in the catalytic mechanism.

-Consensus pattern: [LIVMA]-[AG]-[IVT]-[LIVMFY]-[AG]-x-G-[NHKRQGSAC]-[LIV]-G-x(13,14)-[LIVfMT]-x(2)-[FYwCTH]-[DNSTK]
-Consensus pattern: [LIVMFYWA]-[LIVFYWC]-x(2)-[SAC]-[DNQHR]-[IVFA]-[LIVF]-x-[LIVF]-[HNI]-x-P-x(4)-[STN]-x(2)-[LIVMF]-x-[GSDN]
-Consensus pattern: [LMFATC]-[KPQ]-x-[GSTDN]-x-[LIVMFYWR]-[LIVMFYW](2)-N-x-[STAGC]-R-[GP]-x-[LIVH]-[LIVMC]-[DNV]

[1] Grant G.A. Biochem. Biophys. Res. Commun. 165:1371-1374(1989).
[2] Kochhar S., Hunziker P., Leong-Morgenthaler P.M., Hottinger H. Biochem. Biophys. Res. Commun. 184:60-66(1992).
[3] Ohta T., Taguchi H. J. Biol. Chem. 266:12588-12594(1991).

[4] Goldberg J.D., Yoshida T., Brick P. J. Mol. Biol. 236:1123-1140(1994).

[5] Popov V.O., Lamzin V.S. Biochem. J. 301:625-643(1994).

734. 2-oxo acid dehydrogenases acyltransferase (catalytic domain)

Refined crystal structure of the catalytic domain of dihysrolipoyl transacetylase (E2P) from azotobacter vineelandii at 2.6 angstroms resolution.

Mattevi A, Obmolova G, Kalk KH, Westphal AH, De Kok A, Hol WG; J Mol Biol 1993;230:1183-1199.

These proteins contain one to three copies of a lipoyl binding domain followed by the catalytic domain.

735. 3-beta hydroxysteriod dehydrogenase/isomerase family

Structure and tissue-specific expression of 3 beta-hydroxysteroid dehydrogenase/5-ene-4-ene isomerase genes in human and rat classical and peripheral steroidogenic tissues.

Labrie F, Simard J, Luu-The V, Pelletier G, Belanger A, Lachance Y, Zhao HF, Labrie C, Breton N, de Launoit Y, et al J Steroid Biochem Mol Biol 1992;41:421-435.

The enzyme 3 beta-hydroxysteroid dehydrogenase/5-ene-4-ene isomerase (3 beta-HSD) catalyzes the oxidation and isomerization of 5-ene-3 beta-hydroxypregnene and 5-ene-hydroxyandrostene steroid precursors into the corresponding 4-ene-ketosteroids necessary for the formation of all classes of steroid hormones.

736. 3-hydroxyacyl-CoA dehydrogenase

This family also includes lambda crystallin.

Structure of L-3-hydroxyacyl-coenzyme A dehydrogenase: preliminary chain tracing at 2.8-A resolution.

Birktoft JJ, Holden HM, Hamlin R, Xuong NH, Banaszak LJ;

Proc Natl Acad Sci U S A 1987;84:8262-8266.

3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) (HCDH) [1] is an enzyme involved in fatty acid metabolism, it catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. Most eukaryotic cells have 2 fatty-acid beta-oxidation systems, one located in mitochondria and the other in peroxisomes. In peroxisomes 3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA hydratase (ECH) and 3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme where the N-terminal domain bears the hydratase/isomerase activities and the C-terminal domain the dehydrogenase activity. There are two mitochondrial enzymes: one which is monofunctional and the other which is, like its peroxisomal counterpart, multifunctional.

In Escherichia coli (gene fadB) and Pseudomonas fragi (gene faoA) HCDH is part of a multifunctional enzyme which also contains an ECH/ECI domain as well as a 3-hydroxybutyryl-CoA epimerase domain [2].

The other proteins structurally related to HCDH are:

- Bacterial 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) which reduces 3-hydroxybutanoyl-CoA to acetoacetyl-CoA [3].
- Eye lens protein lambda-crystallin [4], which is specific to lagomorphes (such as rabbit).

There are two major region of similarities in the sequences of proteins of the HCDH family, the first one located in the N-terminal, corresponds to the NAD-binding site, the second one is located in the center of the sequence. A signature pattern has been derived from this central region.

-Consensus pattern: [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-[PA]-[LIVMFY](2)-x(5)-[LIVMFYCT]-[LIVMFY]-x(2)-[GV]

[1] Birktoff J.J., Holden H.M., Hamlin R., Xuong N.-H., Banaszak L.J.
Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987).

[2] Nakahigashi K., Inokuchi H.
Nucleic Acids Res. 18:4937-4937(1990).
[3] Mullany P., Clayton C.L., Pallen M.J., Slone R., Al-Saleh A., Tabaqchali S.
FEMS Microbiol. Lett. 124:61-67(1994).
[4] Mulders J.W.M., Hendriks W., Blankesteijn W.M., Bloemendal H., de Jong W.W.
J. Biol. Chem. 263:15462-15466(1988).

737. 60s Acidic ribosomal protein

Proteins P1, P2, and P0, components of the eukaryotic ribosome stalk. New structural and functional aspects.
Remacha M, Jimenez-Diaz A, Santos C, Briones E, Zambrano R, Rodriguez Gabriel MA, Guarinos E, Ballesta JP;
Biochem Cell Biol 1995;73:959-968.

This family includes archaebacterial L12, eukaryotic P0, P1 and P2.

738. 6-phosphogluconate dehydrogenases

6-phosphogluconate dehydrogenase (EC 1.1.1.44) (6PGD) catalyzes the third step in the hexose monophosphate shunt, the decarboxylating reduction of 6-phosphogluconate in to ribulose 5-phosphate.

Prokaryotic and eukaryotic 6PGD are proteins of about 470 amino acids whose sequence are highly conserved [1]. A region which has been shown [2], from studies of the sheep 6PGD tertiary structure, to be involved in the binding of 6-phosphogluconate has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x-D-x(2)-[GA]-[NQS]-K-G-T-G-x-W

[1] Reizer A., Deutscher J., Saier M.H. Jr., Reizer J.
Mol. Microbiol. 5:1081-1089(1991).
[2] Adams M.J., Archibald I.G., Bugg C.E., Carne A., Gover S.,

Helliwell J.R., Pickersgill R.W., White S.W.
EMBO J. 2:1009-1014(1983).

739. (7tm 1) G-protein coupled receptors [1 to 4,E1,E2] (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. The receptors that are currently known to belong to this family are listed below.

- 5-hydroxytryptamine (serotonin) 1A to 1F, 2A to 2C, 4, 5A, 5B, 6 and 7 [5].
- Acetylcholine, muscarinic-type, M1 to M5.
- Adenosine A1, A2A, A2B and A3 [6].
- Adrenergic alpha-1A to -1C; alpha-2A to -2D; beta-1 to -3 [7].
- Angiotensin II types I and II.
- Bombesin subtypes 3 and 4.
- Bradykinin B1 and B2.
- c3a and C5a anaphylatoxin.
- Cannabinoid CB1 and CB2.
- Chemokines C-C CC-CKR-1 to CC-CKR-8.
- Chemokines C-X-C CXC-CKR-1 to CXC-CKR-4.
- Cholecystokinin-A and cholecystokinin-B/gastrin.
- Dopamine D1 to D5 [8].
- Endothelin ET-a and ET-b [9].
- fMet-Leu-Phe (fMLP) (N-formyl peptide).
- Follicle stimulating hormone (FSH-R) [10].
- Galanin.
- Gastrin-releasing peptide (GRP-R).
- Gonadotropin-releasing hormone (GNRH-R).
- Histamine H1 and H2 (gastric receptor I).
- Lutropin-choriogonadotropic hormone (LSH-R) [10].
- Melanocortin MC1R to MC5R.
- Melatonin.
- Neuromedin B (NMB-R).

- Neuromedin K (NK-3R).
- Neuropeptide Y types 1 to 6.
- Neurotensin (NT-R).
- Octopamine (tyramine), from insects.
- Odorants [11].
- Opioids delta-, kappa- and mu-types [12].
- Oxytocin (OT-R).
- Platelet activating factor (PAF-R).
- Prostacyclin.
- Prostaglandin D2.
- Prostaglandin E2, EP1 to EP4 subtypes.
- Prostaglandin F2.
- Purinoreceptors (ATP) [13].
- Somatostatin types 1 to 5.
- Substance-K (NK-2R).
- Substance-P (NK-1R).
- Thrombin.
- Thromboxane A2.
- Thyrotropin (TSH-R) [10].
- Thyrotropin releasing factor (TRH-R).
- Vasopressin V1a, V1b and V2.
- Visual pigments (opsins and rhodopsin) [14].
- Proto-oncogene mas.
- A number of orphan receptors (whose ligand is not known) from mammals and birds.
- Caenorhabditis elegans putative receptors C06G4.5, C38C10.1, C43C3.2, T27D1.3 and ZC84.4.
- Three putative receptors encoded in the genome of cytomegalovirus: US27, US28, and UL33.
- ECRF3, a putative receptor encoded in the genome of herpesvirus saimiri.

The structure of all these receptors is thought to be identical. They have seven hydrophobic regions, each of which most probably spans the membrane. The N-terminus is located on the extracellular side of the membrane and is

often glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. Most, but not all of these receptors, lack a signal peptide. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved acidic-Arg-aromatic triplet is present in the N-terminal extremity of the second cytoplasmic loop [15] and could be implicated in the interaction with G proteins.

To detect this widespread family of proteins, a pattern that contains the conserved triplet and that also spans the major part of the third transmembrane helix has been developed.

-Consensus pattern: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-
[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM]

[1] Strosberg A.D.
Eur. J. Biochem. 196:1-10(1991).
[2] Kerlavage A.R.
Curr. Opin. Struct. Biol. 1:394-401(1991).
[3] Probst W.C., Snyder L.A., Schuster D.I., Brosius J., Sealfon S.C.
DNA Cell Biol. 11:1-20(1992).
[4] Savarese T.M., Fraser C.M.
Biochem. J. 283:1-9(1992).
[5] Branchek T.
Curr. Biol. 3:315-317(1993).
[6] Stiles G.L.
J. Biol. Chem. 267:6451-6454(1992).
[7] Friell T., Kobilka B.K., Lefkowitz R.J., Caron M.G.
Trends Neurosci. 11:321-324(1988).
[8] Stevens C.F.
Curr. Biol. 1:20-22(1991).

[9] Sakurai T., Yanagisawa M., Masaki T.
Trends Pharmacol. Sci. 13:103-107(1992).
[10] Salesse R., Remy J.J., Levin J.M., Jallal B., Garnier J.
Biochimie 73:109-120(1991).
[11] Lancet D., Ben-Arie N.
Curr. Biol. 3:668-674(1993).
[12] Uhl G.R., Childers S., Pasternak G.
Trends Neurosci. 17:89-93(1994).
[13] Barnard E.A., Burnstock G., Webb T.E.
Trends Pharmacol. Sci. 15:67-70(1994).
[14] Applebury M.L., Hargrave P.A.
Vision Res. 26:1881-1895(1986).
[15] Attwood T.K., Eliopoulos E.E., Findlay J.B.C.
Gene 98:153-159(1991).

(7tm 1) Visual pigments (opsins) retinal binding site

Visual pigments [1,2] are the light-absorbing molecules that mediate vision. They consist of an apoprotein, opsin, covalently linked to the chromophore cis-retinal. Vision is effected through the absorption of a photon by cis-retinal which is isomerized to trans-retinal. This isomerization leads to a change of conformation of the protein. Opsins are integral membrane proteins with seven transmembrane regions that belong to family 1 of G-protein coupled receptors.

In vertebrates four different pigments are generally found. Rod cells, which mediate vision in dim light, contain the pigment rhodopsin. Cone cells, which function in bright light, are responsible for color vision and contain three or more color pigments (for example, in mammals: red, blue and green).

In Drosophila, the eye is composed of 800 facets or ommatidia. Each ommatidium contains eight photoreceptor cells (R1-R8): the R1 to R6 cells are outer cells, R7 and R8 inner cells. Each of the three types of cells (R1-R6, R7 and R8) expresses a specific opsin.

Proteins evolutionary related to opsins include squid retinochrome, also known as retinal photoisomerase, which converts various isomers of retinal into 11-cis retinal and mammalian retinal pigment epithelium (RPE) RGR [3], a protein that may also act in retinal isomerization.

The attachment site for retinal in the above proteins is a conserved lysine residue in the middle of the seventh transmembrane helix. The pattern that had been developed includes this residue.

-Consensus pattern: [LIVMWAC]-[PGC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-[STACP]-
x(2)-[DENF]-[AP]-x(2)-[IY]
[K is the retinal binding site]

[1] Applebury M.L., Hargrave P.A.
Vision Res. 26:1881-1895(1986).
[2] Fryxell K.J., Meyerowitz E.M.
J. Mol. Evol. 33:367-378(1991).
[3] Shen D., Jiang M., Hao W., Tao L., Salazar M., Fong H.K.W.
Biochemistry 33:13117-13125(1994).

The following descriptions of protein family functions are not provided by the Pfam or Prosite databases.

740. BAH

BAH domain. Number of members: 65

[1] Medline: 97074677. Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain. Nicolas RH, Goodwin GH; Gene 1996;175:233-240.
[2] Medline: 99198739. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. Callebaut I, Courvalin J-C, Mornon JP; FEBS letts 1999;446:189-193.

741. ELM2.

ELM2 domain. The ELM2 (Egl-27 and MTA1 homology 2) domain is a small domain of unknown function. Number of members: 10

742. Euk proin. EUKARYOTIC_PORIN The major protein of the outer mitochondrial membrane of eukaryotes is a porin that forms a voltage-dependent anion-selective channel (VDAC) that behaves as a general diffusion pore for small hydrophilic molecules [1 to 4]. The channel adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV.

This protein contains about 280 amino acids and its sequence is composed of between 12 to 16 beta-strands that span the mitochondrial outer membrane. Yeast contains two members of this family (genes POR1 and POR2); vertebrates have at least three members (genes VDAC1, VDAC2 and VDAC3) [5].

A conserved region located at the C-terminal part of these proteins was selected as a signature pattern.

Consensus pattern[YH]-x(2)-D-[SPCAD]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]-[DNSTA]-[DNS]-x(4)-[GSTAN]-[LIVMA]-x-[LIVMY]

[1] Benz R. Biochim. Biophys. Acta 1197:167-196(1994).

[2] Manella C.A. Trends Biochem. Sci. 17:315-320(1992).

[3] Dihanich M. Experientia 46:146-153(1990).

[4] Forte M., Guy H.R., Mannella C.A. J. Bioenerg. Biomembr. 19:341-350(1987).

[5] Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J. Genomics 36:192-196(1996).

743. Glyco hydor 19

Chitinases family 19 signatures

cross-reference(s) CHITINASE_19_1, CHITINASE_19_2

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence

similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19 (also known as classes IA or I and IB or II) are enzymes from plants that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (see the relevant entry <PDOC00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues.

Two highly conserved regions were selected as signature patterns, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond.

Consensus patternC-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA]

Consensus pattern[LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM]

[1]Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992).

[2] Henrissat B. Biochem. J. 280:309-316(1991).

744. MBD

Methyl-CpG binding domain

The Methyl-CpG binding domain (MBD) binds to DNA that contains one or more symmetrically methylated CpGs [1]. DNA methylation in animals is associated with alterations in chromatin structure and silencing of gene expression. MBD has negligible non-specific affinity for DNA. In vitro foot-printing with MeCP2 showed the MBD can protect a 12 nucleotide region surrounding a methyl CpG pair [1]. MBDs are found in several Methyl-CpG binding proteins and also DNA demethylase [2]. Number of members: 11

[1]Medline: 94232813. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nan X, Meehan RR, Bird A; Nucleic Acids Res 1993;21:4886-4892.

[2]Medline: 99158138. A mammalian protein with specific demethylase activity for mCpG DNA. Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M; Nature 1999;397:579-583.

745. Peptidase C1

Eukaryotic thiol (cysteine) proteases active sites

cross-reference(s) THIOL_PROTEASE_CYS; THIOL_PROTEASE_HIS; THIOL_PROTEASE_ASN

Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences).

- Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2].
- Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2].
- Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium- activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain.
- Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3].
- Human cathepsin O [4].
- Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide).
- Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A.
- House-dust mites allergens DerP1 and EurM1.
- Cathepsin B-like proteinases from the worms Caenorhabditis elegans (genes gcp-1, cpr-3, cpr-4, cpr-5 and cpr-6), Schistosoma mansoni (antigen SM31) and Japonica (antigen SJ31), Haemonchus contortus (genes AC-1 and AC-2), and Ostertagia ostertagi (CP-1 and CP-3).
- Slime mold cysteine proteinases CP1 and CP2.
- Cruzipain from Trypanosoma cruzi and brucei.
- Throphozoite cysteine proteinase (TCP) from various Plasmodium species.
- Proteases from Leishmania mexicana, Theileria annulata and Theileria parva.

- Baculoviruses cathepsin-like enzyme (v-cath).
- Drosophila small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain.
- Yeast thiol protease BLH1/YCP1/LAP3.
- Caenorhabditis elegans hypothetical protein C06G4.2, a calpain-like protein.

Two bacterial peptidases are also part of this family:

- Aminopeptidase C from Lactococcus lactis (gene pepC) [5].
- Thiol protease tpr from Porphyromonas gingivalis.

Three other proteins are structurally related to this family, but may have lost their proteolytic activity.

- Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine.
- Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see <PDOC00382>).
- Plasmodium falciparum serine-repeat protein (SERA), the major blood stage antigen. This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine.

The sequences around the three active site residues are well conserved and can be used as signature patterns.

Consensus patternQ-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]

Note the residue in position 4 of the pattern is almost always cysteine; the only exceptions are calpains (Leu), bleomycin hydrolase (Ser) and yeast YCP1 (Ser). Note the residue in position 5 of the pattern is always Gly except in papaya protease IV where it is Glu.

Consensus pattern[LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH] [H is the active site residue]

Consensus pattern[FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue]

Note these proteins belong to family C1 (papain-type) and C2 (calpains) in the classification of peptidases [7,E1].

[1]Dufour E. Biochimie 70:1335-1342(1988).
[2]Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).
[3]Shi G.-P., Chapman H.A., Bhairi S.M., Deleeuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).
[4]Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).
[5]Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).
[6]Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).
[7]Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

746. Peptidase M22

Glycoprotease family signature cross-reference(s) GLYCOPROTEASE

Glycoprotease (GCP) (EC 3.4.24.57) [1], or o-syaloglycoprotein endopeptidase, is a metalloprotease secreted by Pasteurella haemolytica which specifically cleaves O-sialoglycoproteins such as glycophorin A. The sequence of GCP is highly similar to the following uncharacterized proteins:

- Escherichia coli hypothetical protein ygjD (ORF-X).
- Bacillus subtilis hypothetical protein ydiE.
- Mycobacterium leprae hypothetical protein U229E.
- Mycobacterium tuberculosis hypothetical protein MtCY78.10.
- Synechocystis strain PCC 6803 hypothetical protein slr0807.
- Methanococcus jannaschii hypothetical protein MJ1130.
- Haloarcula marismortui hypothetical protein in HSH 3'region.
- Yeast hypothetical protein YKR038c.
- Yeast hypothetical protein QRI7.

One of the conserved regions contains two conserved histidines. It is possible that this region is involved in coordinating a metal ion such as zinc.

Consensus pattern[KR]-[GSAT]-x(4)-[FYWLH]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H-x(2)-[AG]-H-[LIVM]

Note these proteins belong to family M22 in the classification of peptidases [2,E1].

[1]Abdullah K.M., Lo R.Y.C., Mellors A. J. Bacteriol. 173:5597-5603(1991).

[2]Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

747. SAM. SAM domain (Sterile alpha motif)

It has been suggested that SAM is an evolutionarily conserved protein binding domain that is involved in the regulation of numerous developmental processes in diverse eukaryotes. The SAM domain can potentially function as a protein interaction module through its ability to homo- and heterooligomerise with other SAM domains. Number of members: 81

[1]Medline: 96100659 SAM: A novel motif in yeast sterile alpha and Drosophila polyhomeotic proteins Ponting CP; Prot Sci 1995;4:1928-1930.

[2]Medline: 97160498 SAM as a protein interaction domain involved in developmental regulation. Shultz J, Ponting CP, Hofmann K, Bork P; Prot Sci 1997;6:249-253.

[3]Medline: 99101382 The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. Reference Author: Stapleton D, Balan I, Pawson T, Sicheri F; Nat Struct Biol 1999;6:44-49.

748. Tyrosinase signatures cross-reference(s) TYROSINASE_1; TYROSINASE_2

Tyrosinase (EC 1.14.18.1) [1] is a copper monooxygenases that catalyzes the hydroxylation of monophenols and the oxidation of o-diphenols to o-quinols. This enzyme, found in prokaryotes as well as in eukaryotes, is involved in the formation of pigments such as melanins and other polyphenolic compounds.

Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ion has been shown [2] to be bound by three conserved histidines residues. The regions around these copper-binding ligands are well conserved and also shared by some

hemocyanins, which are copper-containing oxygen carriers from the hemolymph of many molluscs and arthropods [3,4].

At least two proteins related to tyrosinase are known to exist in mammals:

- TRP-1 (TYRP1) [5], which is responsible for the conversion of 5,6-dihydro-xyindole-2-carboxylic acid (DHICA) to indole-5,6-quinone-2-carboxylic acid.
- TRP-2 (TYRP2) [6], which is the melanogenic enzyme DOPAchrome tautomerase (EC 5.3.3.12) that catalyzes the conversion of DOPAchrome to DHICA. TRP-2 differs from tyrosinases and TRP-1 in that it binds two zinc ions instead of copper [7].

Other proteins that belong to this family are:

- Plants polyphenol oxidases (PPO) (EC 1.10.3.1) which catalyze the oxidation of mono- and o-diphenols to o-diquinones [8].
- Caenorhabditis elegans hypothetical protein C02C2.1.

Two signature patterns for tyrosinase and related proteins have been derived The first one contains two of the histidines that bind CuA, and is located in the N-terminal section of tyrosinase. The second pattern contains a histidine that binds CuB, that pattern is located in the central section of the enzyme.

Consensus pattern H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LM]-x(3)-E
[The two H's are copper ligands]
Consensus patternD-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D [H is a copper ligand]

[1]Lerch K. Prog. Clin. Biol. Res. 256:85-98(1988).
[2]Jackman M.P., Hajnal A., Lerch K. Biochem. J. 274:707-713(1991).
[3]Linzen B. Naturwissenschaften 76:206-211(1989).
[4]Lang W.H., van Holde K.E. Proc. Natl. Acad. Sci. U.S.A. 88:244-248(1991).
[5]Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borron J.C., Hearing V.J. EMBO J. 13:5818-5825(1994).

[6]Jackson I.J., Chambers D.M., Tsukamoto K., Copeland N.G., Gilbert D.J., Jenkins N.A., Hearing V. EMBO J. 11:527-535(1992).

[7]Solano F., Martinez-Liarte J.H., Jimenez-Cervantes C., Garcia-Borron J.C., Lozano J.A. Biochem. Biophys. Res. Commun. 204:1243-1250(1994).

[8]Cary J.W., Lax A.R., Flurkey W.H. Plant Mol. Biol. 20:245-253(1992).

749. (Mur Ligase) Folylpolyglutamate synthase signatures

Folylpolyglutamate synthase (EC 6.3.2.17) (FPGS) [1] is the enzyme of folate metabolism that catalyzes ATP-dependent addition of glutamate moieties to tetrahydrofolate.

Its sequence is moderately conserved between prokaryotes (gene folC) and eukaryotes. We developed two signature patterns based on the conserved regions which are rich in glycine residues and could play a role in the catalytical activity and/or in substrate binding.

Description of pattern(s) and/or profile(s)

Consensus pattern[LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)- [LIVM](2)-x(3)-[GSK]

Consensus pattern[LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2)

[1]Shane B., Garrow T., Brenner A., Chen L., Choi Y.J., Hsu J.C., Stover P. Adv. Exp. Med. Biol. 338:629-634(1993).

750. (Peptidase M3) Neutral zinc metallopeptidases, zinc-binding region signature

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily,known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

Family M1

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
- Mammalian aminopeptidase N (EC 3.4.11.2).
- Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
- Yeast aminopeptidase yscII (gene APE2).
- Yeast alanine/arginine aminopeptidase (gene AAP1).
- Yeast hypothetical protein YIL137c.
- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

Family M2

- Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.

Family M3

- Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.
- Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).
- Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
- Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).
- Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).
- Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).
- Yeast hypothetical protein YKL134c.

Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus.
- Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB).
- Extracellular elastase from Staphylococcus epidermidis.

- Extracellular protease prt1 from Erwinia carotovora.
- Extracellular minor protease smp from Serratia marcescens.
- Vibriolysin (EC 3.4.24.25) from various species of Vibrio.
- Protease prtA from Listeria monocytogenes.
- Extracellular proteinase proA from Legionella pneumophila.

Family M5

- Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.

Family M6

- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

Family M7

- Streptomyces extracellular small neutral proteases

Family M8

- Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of Leishmania.

Family M9

- Microbial collagenase (EC 3.4.24.3) from Clostridium perfringens and Vibrio alginolyticus.

Family M10A

- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from Serratia.
- Alkaline metalloproteinase from Pseudomonas aeruginosa (gene aprA).
- Secreted proteases A, B, C and G from Erwinia chrysanthemi.
- Yeast hypothetical protein YIL108w.

Family M10B

- Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34)

(neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).

- Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.

- Soybean metalloendoproteinase 1.

Family M11

- Chlamydomonas reinhardtii gamete lytic enzyme (GLE).

Family M12A

- Astacin (EC 3.4.24.21), a crayfish endoprotease.

- Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border metalloendopeptidase.

- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The Drosophila homolog of BMP-1 is the dorsal-ventral patterning protein tolloid.

- Blastula protease 10 (BP10) from Paracentrotus lividus and the related protein SpAN from Strongylocentrotus purpuratus.

- Caenorhabditis elegans protein toh-2.

- Caenorhabditis elegans hypothetical protein F42A10.8.

- Choriolysins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish Oryzias lapides. These proteases participates in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.

Family M12B

- Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53).

- Mouse cell surface antigen MS2.

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).
- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.
- Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.
- Peptidase O from Lactococcus lactis (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

Family M30

- Staphylococcus hyicus neutral metalloprotease.

Family M32

- Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from Thermus aquaticus which is most active at high temperature.

Family M34

- Lethal factor (LF) from Bacillus anthracis, one of the three proteins composing the anthrax toxin.

Family M35

- Deuterolysin (EC 3.4.24.39) from Penicillium citrinum and related proteases from various species of Aspergillus.

Family M36

- Extracellular elastinolytic metalloproteinases from Aspergillus.

From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is

a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Description of pattern(s) and/or profile(s)

Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ] [The

two H's are zinc ligands] [E is the active site residue]

Sequences known to belong to this class detected by the patternALL, except for members of families M5, M7 amd M11.

Other sequence(s) detected in SWISS-PROT55; including Neurospora crassa conidiation-specific protein 13 which could be a zinc-protease.

[1]Jongeneel C.V., Bouvier J., Bairoch A.

FEBS Lett. 242:211-214(1989).

[2]Murphy G.J.P., Murphy G., Reynolds J.J.

FEBS Lett. 289:4-7(1991).

[3]Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B., Stoecker W.

Zoology 99:237-246(1996).

[4]Rawlings N.D., Barrett A.J.

Meth. Enzymol. 248:183-228(1995).

[5]Woessner J. Jr.

FASEB J. 5:2145-2154(1991).

[6]Hite L.A., Fox J.W., Bjarnason J.B.

[7]Montecucco C., Schiavo G.

Trends Biochem. Sci. 18:324-327(1993).

[8]Niemann H., Blasi J., Jahn R.

Trends Cell Biol. 4:179-185(1994).

751. PseudoU_synt_1

tRNA pseudouridine synthase is involved in the formation of pseudouridine at the anticodon stem and loop of transfer-RNAs Pseudouridine is an isomer of uridine (5-(beta-D-ribofuranosyl) uracil, and id the most abundant modified nucleoside found in all cellular RNAs. The TruA-like proteins also exhibit a conserved sequence with a strictly conserved aspartic acid, likely involved in catalysis. Number of members: 25

[1]Medline: 98254513. Transfer RNA-pseudouridine synthetase Pus1 of Saccaromyces cerevisiae contains one atom of zinc essential for its native conformation and tRNA recognition. Arluison V, Hountondji C, Robert B, Grosjean H; Biochemistry 1998;37:7268-7276.

752. EPSP synthase signatures

EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase) (EC 2.5.1.19) catalyzes the sixth step in the biosynthesis from chorismate of the aromatic amino acids (the shikimate pathway) in bacteria (gene aroA), plants and fungi (where it is part of a multifunctional enzyme which catalyzes five consecutive steps in this pathway) [1]. EPSP synthase has been extensively studied as it is the target of the potent herbicide glyphosate which inhibits the enzyme.

The sequence of EPSP from various biological sources shows that the structure of the enzyme has been well conserved throughout evolution. Two conserved regions were selected as signature patterns. The first pattern corresponds to a region that is part of the active site and which is also important for the resistance to glyphosate [2]. The second pattern is located in the C-terminal part of the protein and contains a conserved lysine which seems to be important for the activity of the enzyme.

Description of pattern(s) and/or profile(s)

Consensus pattern[LIVM]-x(2)-[GN]-N-[SA]-G-T-[STA]-x-R-x-[LIVMY]-x-[GSTA]
Consensus pattern[KR]-x-[KH]-E-[CST]-[DNE]-R-[LIVM]-x-[STA]-[LIVMC]-x(2)-[EN]-[LIVMF]-x-[KRA]-[LIVMF]-G

[1]Stallings W.C., Abdel-Megid S.S., Lim L.W., Shieh H.-S., Dayringer H.E., Leimgruber N.K., Stegeman R.A., Anderson K.S., Sikorski J.A., Padgette S.R., Kishore G.M. Proc. Natl. Acad. Sci. U.S.A. 88:5046-5050(1991).

[2]Padgette S.R., Re D.B., Gaser C.S., Eicholtz D.A., Frazier R.B., Hironaka C.M., Levine E.B., Shah D.M., Fraley R.T., Kishore G.M. J. Biol. Chem. 266:22364-22369(1991).

753. Glyco_hydro_18

Glycosyl hydrolases family 18. Number of members: 173

[1]Medline: 95219379. Crystal structure of a bacterial chitinase at 2.3 A resolution. Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS, Vorgias CE; Structure 1994;2:1169-1180.

754. Esterase

Putative esterase

This family contains Esterase D Swiss:P10768. However it is not clear if all members of the family have the same function. This family is possibly related to the COesterase family. Number of members: 36

755. (HMA) Heavy-metal-associated domain

A conserved domain of about 30 amino acid residues has been found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in:

- A variety of cation transport ATPases (E1-E2 ATPases) (see <PDOC00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pI258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases

that contain the HMA domain are: fixI from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal section.

- Mercuric reductase (EC 1.16.1.1) (gene merA) which is generally encoded by plasmids carried by mercury-resistant Gram-negative bacteria. Mercuric reductase is a class-1 pyridine nucleotide-disulphide oxidoreductase (see <PDOC00073>). There is generally one HMA domain (with the exception of a chromosomal merA from Bacillus strain RC607 which has two) in the N-terminal part of merA.
- Mercuric transport protein periplasmic component (gene merP), also encoded by plasmids carried by mercury-resistant Gram-negative bacteria. It seems to be a mercury scavenger that specifically binds to one Hg(2+) ion and which passes it to the mercuric reductase via the merT protein. The N-terminal half of merP is a HMA domain.
- Helicobacter pylori copper-binding protein copP.
- Yeast protein ATX1 [2], which could act in the transport and/or partitioning of copper.

The consensus pattern for HMA spans the complete domain.

Description of pattern(s) and/or profile(s)
Consensus pattern[LIVN]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS] [The two C's probably bind metals]

[1]Bull P.C., Cox D.W. Trends Genet. 10:246-252(1994).
[2]Lin S.-J., Culotta V.L. Proc. Natl. Acad. Sci. U.S.A. 92:3784-3788(1995).

756. (Peptidase M10) Matrixins cysteine switch
PROSITE cross-reference(s): CYSTEINE_SWITCH
Mammalian extracellular matrix metalloproteinases (EC 3.4.24.-), also known as matrixins [1] (see <PDOC00129>), are zinc-dependent enzymes. They are secreted by cells in an inactive form (zymogen) that differs from the mature enzyme by the presence of an N-terminal propeptide. A highly conserved octapeptide is found two residues downstream of the C-terminal end of the propeptide. This region has been shown to be involved in

autoinhibition of matrixins [2,3]; a cysteine within the octapeptide chelates the active site zinc ion, thus inhibiting the enzyme. This region has been called the 'cysteine switch' or 'autoinhibitor region'.

A cysteine switch has been found in the following zinc proteases:

- MMP-1 (EC 3.4.24.7) (interstitial collagenase).
- MMP-2 (EC 3.4.24.24) (72 Kd gelatinase).
- MMP-3 (EC 3.4.24.17) (stromelysin-1).
- MMP-7 (EC 3.4.24.23) (matrilysin).
- MMP-8 (EC 3.4.24.34) (neutrophil collagenase).
- MMP-9 (EC 3.4.24.35) (92 Kd gelatinase).
- MMP-10 (EC 3.4.24.22) (stromelysin-2).
- MMP-11 (EC 3.4.24.-) (stromelysin-3).
- MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).
- MMP-13 (EC 3.4.24.-) (collagenase 3).
- MMP-14 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 1).
- MMP-15 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 2).
- MMP-16 (EC 3.4.24.-) (membrane-type matrix metalliproteinase 3).
- Sea urchin hatching enzyme (EC 3.4.24.12) (envelysin) [4].
- Chlamydomonas reinhardtii gamete lytic enzyme (GLE) [5].

Description of pattern(s) and/or profile(s)

Consensus patternP-R-C-[GN]-x-P-[DR]-[LIVSAPKQ] [C chelates the zinc ion]

[1]Woessner J. Jr. FASEB J. 5:2145-2154(1991).

[2]Sanchez-Lopez R., Nicholson R., Gesnel M.C., Matrisian L.M., Breathnach R. J. Biol. Chem. 263:11892-11899(1988).

[3]Park A.J., Matrisian L.M., Kells A.F., Pearson R., Yuan Z., Navre M. J. Biol. Chem. 266:1584-1590(1991).

[4]Lepage T., Gache C. EMBO J. 9:3003-3012(1990).

[5]Kinoshita T., Fukuzawa H., Shimada T., Saito T., Matsuda Y. Proc. Natl. Acad. Sci. U.S.A. 89:4693-4697(1992).

757. (Peptidase S8) Serine proteases, subtilase family, active sites

PROSITE cross-reference(s): PS00136; SUBTILASE_ASP, PS00137; SUBTILASE_HIS, PS00138; SUBTILASE_SER

Subtilases [1,2] are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases and can be used as signatures specific to that category of proteases.

The subtilase family currently includes the following proteases:

- Subtilisins (EC 3.4.21.62), these alkaline proteases from various Bacillus species have been the target of numerous studies in the past thirty years.
- Alkaline elastase YaB from Bacillus sp. (gene ale).
- Alkaline serine exoprotease A from Vibrio alginolyticus (gene proA).
- Aqualysin I from Thermus aquaticus (gene pstI).
- AspA from Aeromonas salmonicida.
- Bacillopeptidase F (esterase) from Bacillus subtilis (gene bpf).
- C5A peptidase from Streptococcus pyogenes (gene scpA).
- Cell envelope-located proteases PI, PII, and PIII from Lactococcus lactis.
- Extracellular serine protease from Serratia marcescens.
- Extracellular protease from Xanthomonas campestris.
- Intracellular serine protease (ISP) from various Bacillus.
- Minor extracellular serine protease epr from Bacillus subtilis (gene epr).
- Minor extracellular serine protease vpr from Bacillus subtilis (gene vpr).
- Nisin leader peptide processing protease nisP from Lactococcus lactis.
- Serotype-specific antigene 1 from Pasteurella haemolytica (gene ssa1).
- Thermitase (EC 3.4.21.66) from Thermoactinomyces vulgaris.
- Calcium-dependent protease from Anabaena variabilis (gene prcA).
- Halolysin from halophilic bacteria sp. 172p1 (gene hly).
- Alkaline extracellular protease (AEP) from Yarrowia lipolytica (gene xpr2).
- Alkaline proteinase from Cephalosporium acremonium (gene alp).
- Cerevisin (EC 3.4.21.48) (vacuolar protease B) from yeast (gene PRB1).
- Cuticle-degrading protease (pr1) from Metarhizium anisopliae.
- KEX-1 protease from Kluyveromyces lactis.

- Kexin (EC 3.4.21.61) from yeast (gene KEX-2).
- Oryzin (EC 3.4.21.63) (alkaline proteinase) from Aspergillus (gene alp).
- Proteinase K (EC 3.4.21.64) from Tritirachium album (gene proK).
- Proteinase R from Tritirachium album (gene proR).
- Proteinase T from Tritirachium album (gene proT).
- Subtilisin-like protease III from yeast (gene YSP3).
- Thermomycolin (EC 3.4.21.65) from Malbranchea sulfurea.
- Furin (EC 3.4.21.85), neuroendocrine convertases 1 to 3 (NEC-1 to -3) and PACE4 protease from mammals, other vertebrates, and invertebrates. These proteases are involved in the processing of hormone precursors at sites comprised of pairs of basic amino acid residues [3].
- Tripeptidyl-peptidase II (EC 3.4.14.10) (tripeptidyl aminopeptidase) from Human.
- Prestalk-specific proteins tagB and tagC from slime mold [4]. Both proteins consist of two domains: a N-terminal subtilase catalytic domain and a C-terminal ABC transporter domain (see <PDOC00185>).

Description of pattern(s) and/or profile(s)

Consensus pattern[STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH] [D is the active site residue]

Consensus patternH-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM] [H is the active site residue]

Consensus patternG-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG] [S is the active site residue]

Note if a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100%

Note these proteins belong to family S8 in the classification of peptidases [5,E1].

[1]Siezen R.J., de Vos W.M., Leunissen J.A.M., Dijkstra B.W. Protein Eng. 4:719-737(1991).

[2]Siezen R.J. (In) Proceeding subtilisin symposium, Hamburg, (1992).

[3]Barr P.J. Cell 66:1-3(1991).

[4]Shaulsky G., Kuspa A., Loomis W.F.; Genes Dev. 9:1111-1122(1995).

[5]Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

758. (SSB) Single-strand binding protein family signatures

PROSITE cross-reference(s): PS00735; SSB_1,PS00736; SSB_2

The Escherichia coli single-strand binding protein [1] (gene ssb), also known as the helix-destabilizing protein, is a protein of 177 amino acids. It binds tightly, as a homotetramer, to single-stranded DNA (ss-DNA) and plays an important role in DNA replication, recombination and repair.

Closely related variants of SSB are encoded in the genome of a variety of large self-transmissible plasmids. SSB has also been characterized in bacteria such as Proteus mirabilis or Serratia marcescens.

Eukaryotic mitochondrial proteins that bind ss-DNA and are probably involved in mitochondrial DNA replication are structurally and evolutionary related to prokaryotic SSB. Proteins currently known to belong to this subfamily are listed below [2].

- Mammalian protein Mt-SSB (P16).
- Xenopus Mt-SSBs and Mt-SSBr.
- Drosophila MtSSB.
- Yeast protein RIM1.

Two signature patterns have been developed for these proteins. The first is a conserved region in the N-terminal section of the SSB's. The second is a centrally located region which, in Escherichia coli SSB, is known to be involved in the binding of DNA.

Description of pattern(s) and/or profile(s)

Consensus pattern[LIVMF]-[NST]-[KRT]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVM]-[GST]-x-[DET]

Consensus patternT-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR]

[1]Meyer R.R., Laine P.S. Microbiol. Rev. 54:342-380(1990).

[2]Stroumbakis N.D., Li Z., Tolias P.P. Gene 143:171-177(1994).

759. KDPG and KHG aldolases active site signatures

PROSITE cross-reference(s): PS00159; ALDOLASE_KDPG_KHG_1, PS00160; ALDOLASE_KDPG_KHG_2

4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) (KHG-aldolase) catalyzes the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate. Phospho-2-dehydro-3-deoxygluconate aldolase (EC 4.1.2.14) (KDPG-aldolase) catalyzes the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde 3-phosphate.

These two enzymes are structurally and functionally related [1]. They are both homotrimeric proteins of approximately 220 amino-acid residues. They are class I aldolases whose catalytic mechanism involves the formation of a Schiff-base intermediate between the substrate and the epsilon-amino group of a lysine residue. In both enzymes, an arginine is required for catalytic activity.

Two signature patterns were developed for these enzymes. The first one contains the active site arginine and the second, the lysine involved in the Schiff-base formation.

Description of pattern(s) and/or profile(s)
Consensus patternG-[LIVM]-x(3)-E-[LIV]-T-[LF]-R [R is the active site residue]
Consensus patternG-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G [K is involved in Schiff-base formation]

[1] Vlahos C J., Dekker E.E. J. Biol. Chem. 263:11683-11691(1988).

760. AP endonucleases family 1 signatures. PROSITE cross-reference(s): PS00726; AP_NUCLEASE_F1_1, PS00727; AP_NUCLEASE_F1_2, PS00728; AP_NUCLEASE_F1_3

DNA damaging agents such as the antitumor drugs bleomycin and neocarzinostatin or those that generate oxygen radicals produce a variety of lesions in DNA. Amongst these is base-loss which forms apurinic/apyrimidinic (AP) sites or strand breaks with atypical 3'termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of the

phosphodiester backbone. Such endonucleases are also generally capable of removing blocking groups from the 3'terminus of DNA strand breaks.

AP endonucleases can be classified into two families on the basis of sequence similarity. Family 1 groups the enzymes listed below [1].

- Escherichia coli exonuclease III (EC 3.1.11.2) (gene xthA).
- Streptococcus pneumoniae and Bacillus subtilis exonuclease A (gene exoA).
- Mammalian AP endonuclease 1 (AP1) (EC 4.2.99.18).
- Drosophila recombination repair protein 1 (gene Rrp1).
- Arabidopsis thaliana apurinic endonuclease-redox protein (gene arp).

Except for Rrp1 and arp, these enzymes are proteins of about 300 amino-acid residues. Rrp1 and arp both contain additional and unrelated sequences in their N-terminal section (about 400 residues for Rrp1 and 270 for arp).

Three signature patterns were developed for this family of enzymes. The patterns are based on the most conserved regions. The first pattern contains a glutamate which has been shown [2], in the Escherichia coli enzyme to bind a divalent metal ion such as magnesium or manganese

Consensus pattern[APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K [E binds a divalent metal ion]
Consensus patternD-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2)
Consensus patternN-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S

[1] Barzilay G., Hickson I.S. BioEssays 17:713-719(1995).
[2] Mol C.D., Kuo C.-F., Thayer M.M., Cunningham R.P., Tainer J.A. Nature 374:381-386(1995).

761. (ER)Enhancer of rudimentary signature, PROSITE cross-reference(s): PS01290; ER

The Drosophila protein 'enhancer of rudimentary' (gene (e(r)) is a small protein of 104 residues whose function is not yet clear. From an evolutionary point of view, it is highly

conserved [1] and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein plays a role in the cell cycle.

A conserved region in the central part of the protein was selected as as signaure pattern.

Consensus patternY-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S

[1] Gelsthorpe M., Pulumati M., McCallum C., Dang-Vu K., Tsubota S.I. Gene 186:189-195(1997).

762. (ETF alpha) Electron transfer flavoprotein alpha-subunit signature, PROSITE cross-reference(s): PS00696; ETF_ALPHA

The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria.

The alpha subunit of ETF is a protein of about 32 Kd which is structurally related to the bacterial nitrogen fixation protein fixB which could play a role in a redox process and feed electrons to ferredoxin.

Other related proteins are:

- Escherichia coli hypothetical protein ydiR.
- Escherichia coli hypothetical protein ygcQ.

A highly conserved region which is located in the C-terminal section was selected as a signature pattern for these proteins.

Consensus pattern [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)-[IV]-x-A-[IV]-N

[1] Finocchiaro G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990).

[2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).

763. (lectin c) C-type lectin domain signature and profile

PROSITE cross-reference(s): PS00615; C_TYPE_LECTIN_1, PS50041; C_TYPE_LECTIN_2

A number of different families of proteins share a conserved domain which was first characterized in some animal lectins and which seem to function as a calcium-dependent carbohydrate-recognition domain [1,2,3]. This domain, which is known as the C-type lectin domain (CTL) or as the carbohydrate-recognition domain (CRD), consists of about 110 to 130 residues. There are four cysteines which are perfectly conserved and involved in two disulfide bonds. A schematic representation of the CTL domain is shown below.

```
    +------+
    |      |
xcxxxxcxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxCxxxxWxCxxxxCx
|    |        |************|*
+----+        +------------------------------------------------+
```

'C': conserved cysteine involved in a disulfide bond.

'c': optional cysteine involved in a disulfide bond.

'*': position of the pattern.

The categories of proteins, in which the CTL domain has been found, are listed below.

Type-II membrane proteins where the CTL domain is located at the C-terminal extremity of the proteins:

- Asialoglycoprotein receptors (ASGPR) (also known as hepatic lectins) [4]. The ASGPR's mediate the endocytosis of plasma glycoproteins to which the terminal sialic acid residue in their carbohydrate moieties has been removed.
- Low affinity immunoglobulin epsilon Fc receptor (lymphocyte IgE receptor), which plays an essential role in the regulation of IgE production and in the differentiation of B cells.

- Kupffer cell receptor. A receptor with an affinity for galactose and fucose, that could be involved in endocytosis.
- A number of proteins expressed on the surface of natural killer T-cells: NKG2, NKR-P1, YE1/88 (Ly-49), CD69 and on B-cells: CD72, LyB-2. The CTL- domain in these proteins is distantly related to other CTL-domains; it is unclear whether they are likely to bind carbohydrates.

Proteins that consist of an N-terminal collagenous domain followed by a CTL- domain [5], these proteins are sometimes called 'collectins':

- Pulmonary surfactant-associated protein A (SP-A). SP-A is a calcium-dependent protein that binds to surfactant phospholipids and contributes to lower the surface tension at the air-liquid interface in the alveoli of the mammalian lung.
- Pulmonary surfactant-associated protein D (SP-D).
- Conglutinin, a calcium-dependent lectin-like protein which binds to a yeast cell wall extract and to immune complexes through the complement component (iC3b).
- Mannan-binding proteins (MBP) (also known as mannose-binding proteins). MBP's bind mannose and N-acetyl-D-glucosamine in a calcium-dependent manner.
- Bovine collectin-43 (CL-43).

Selectins (or LEC-CAM) [6,7]. Selectins are cell adhesion molecules implicated in the interaction of leukocytes with platelets or vascular endothelium. Structurally, selectins consist of a long extracellular domain, followed by a transmembrane region and a short cytoplasmic domain. The extracellular domain is itself composed of a CTL-domain, followed by an EGF-like domain and a variable number of SCR/Sushi repeats. Known selectins are:

- Lymph node homing receptor (also known as L-selectin, leukocyte adhesion molecule-1, (LAM-1), leu-8, gp90-mel, or LECAM-1)
- Endothelial leukocyte adhesion molecule 1 (ELAM-1, E-selectin or LECAM-2). The ligand recognized by ELAM-1 is sialyl-Lewis x.

- Granule membrane protein 140 (GMP-140, P-selectin, PADGEM, CD62, or LECAM-3). The ligand recognized by GMP-140 is Lewis x.

Large proteoglycans that contain a CTL-domain followed by one copy of a SCR/ Sushi repeat, in their C-terminal section:

- Aggrecan (cartilage-specific proteoglycan core protein). This proteoglycan is a major component of the extracellular matrix of cartilagenous tissues where it has a role in the resistance to compression.
- Brevican.
- Neurocan.
- Versican (large fibroblast proteoglycan), a large chondroitin sulfate proteoglycan that may play a role in intercellular signalling.

In addition to the CTL and Sushi domains, these proteins also contain, in their N-terminal domain, an Ig-like V-type region, two or four link domains (see <PDOC00955>) and up to two EGF-like repeats.

Two type-I membrane proteins:

- Mannose receptor from macrophages. This protein mediates the endocytosis of glycoproteins by macrophages in several recognition and uptake processes. Its extracellular section consists of a fibronectin type II domain followed by eight tandem repeats of the CTL domain.
- 180 Kd secretory phospholipase A2 receptor (PLA2-R). A protein whose structure is highly similar to that of the mannose receptor.
- DEC-205 receptor. This protein is used by dendritic cells and thymic epithelial cells to capture and endocytose diverse carbohydrate-binding antigens and direct them to antigen-processing cellular compartiments. DEC-205 extracellular section consists of a fibronectin type II domain followed by ten tandem repeats of the CTL domain.
- Silk moth hemocytin, an humoral lectin which is involved in a self-defence mechanism. It is composed of 2 FA58C domains (see <PDOC00988>), a CTL domain, 2 VWFC domains (see <PDOC00928), and a CTCK (see <PDOC00912>).

Various other proteins that uniquely consist of a CTL domain:

- Invertebrate soluble galactose-binding lectins. A category to which belong a humoral lectin from a flesh fly; echinoidin, a lectin from the coelomic fluid of a sea urchin; BRA-2 and BRA-3, two lectins from the coelomic fluid of a barnacle, a lectin from the tunicate Polyandrocarpa misakiensis and a newt oviduct lectin. The physiological importance of these lectins is not yet known but they may play an important role in defense mechanisms.
- Pancreatic stone protein (PSP) (also known as pancreatic thread protein (PTP), or reg), a protein that might act as an inhibitor of spontaneous calcium carbonate precipitation.
- Pancreatitis associated protein (PAP), a protein that might be involved in the control of bacterial proliferation.
- Tetranectin, a plasma protein that binds to plasminogen and to isolated kringle 4.
- Eosinophil granule major basic protein (MBP), a cytotoxic protein.
- A galactose specific lectin from a rattlesnake.
- Two subunits of a coagulation factor IX/factor X-binding protein (IX/X-bp), a snake venom anticoagulant protein which binds with factors IX and X in the presence of calcium.
- Two subunits of a phospholipase A2 inhibitor from the plasma of a snake (PLI-A and PLI-B).
- A lipopolysaccharide-binding protein (LPS-BP) from the hemolymph of a cockroach [8].
- Sea raven antifreeze protein (AFP) [9].

As a signature pattern for this domain, the C-terminal region with its three conserved cysteines was selected.

Consensus patternC-[LIVMFYATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-C [The three C's are involved in disulfide bonds]

Note all CTL domains have five Trp residues before the second Cys,

with the exception of tunicate lectin and cockroach LPS-BP which have Leu.

Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

[1] Drickamer K. J. Biol. Chem. 263:9557-9560(1988).
[2] Drickamer K. Prog. Nucleic Acid Res. Mol. Biol. 45:207-232(1993).
[3] Drickamer K. Curr. Opin. Struct. Biol. 3:393-400(1993).
[4] Spiess M. Biochemistry 29:10009-10018(1990).
[5] Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A. Science 254:1608-1615(1991).
[6] Siegelman M. Curr. Biol. 1:125-128(1991).
[7] Lasky L.A. Science 238:964-969(1992).
[8] Jomori T., Natori S. J. Biol. Chem. 266:13318-13323(1991).
[9] Ng N.F.L., Hew C.-L. J. Biol. Chem. 267:16069-16075(1992).

764. (SRCR) Speract receptor repeated domain signature
PROSITE cross-reference(s): PS00420; SPERACT_RECEPTOR,

The receptor for the sea urchin egg peptide speract is a transmembrane glycoprotein of 500 amino acid residues [1]. Structurally it consists of a large extracellular domain of 450 residues, followed by a transmembrane region and a small cytoplasmic domain of 12 amino acids. The extracellular domain contains four repeats of a 115 amino acids domain. There are 17 positions that are perfectly conserved in the four repeats, among them are six cysteines, six glycines, and three glutamates.

Such a domain is also found, once, in the C-terminal section of mammalian macrophage scavenger receptor type I [2], a membrane glycoproteins implicated in the pathologic deposition of cholesterol in arterial walls during atherogenesis.

The signature pattern that was derived spans part of the N-terminal section of the domain and contains 8 of the 17 conserved residues.

Consensus patternG-x(5)-G-x(2)-E-x(6)-W-G-x(2)-C-x(3)-[FYW]-x(8)-C-x(3)-G

[1] Dangott J.J., Jordan J.E., Bellet R.A., Garbers D.L. Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).

[2] Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G., Jenkins N.A., Krieger M. Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).

765. Bac_surface_Ag

Bacterial surface antigen

This entry includes the following surface antigens; D15 antigen from H.influenzae, OMA87 from P.multocida, OMP85 from N.meningitidis and N.gonorrhoeae. Number of members: 14

[1]Medline: 95255676. The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae. Flack FS, Loosmore S, Chong P, Thomas WR; Gene 1995;156:97-99.

[2] Medline: 96333354. Cloning, sequencing, expression, and protective capacity of the oma87 gene encoding the Pasteurella multocida 87-kilodalton outer membrane antigen. Ruffolo CG, Adler B; Infect Immun 1996;64:3161-3167.

766. BRCA1 C Terminus (BRCT) domain

The BRCT domain is found predominantly in proteins involved in cell cycle checkpoint functions responsive to DNA damage. It has been suggested that the Retinoblastoma protein contains a divergent BRCT domain, this has not been included in this family. The BRCT domain of XRCC1 forms a homodimer in the crystal structure Medline:99016060. This suggests that pairs of BRCT domains associate as homo- or heterodimers. Number of members: 131

[1] Medline: 96259550. BRCA1 protein products ...Functional motifs... Koonin EV, Altschul SF, Bork P; Nature Genet 1996;13:266-268.

[2] Medline: 97153217. From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair Callebaut I, Mornon JP; Febs lett 1997;400:25-30.

[3] Medline: 97186552. A superfamily of conserved domains in DNA damage responsive cell cycle checkpoint proteins Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV; Faseb J 1997;11:68-76.

[4] Medline: 97402527. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ; Nucleic Acids Res 1997;25:3389-3402.

[5] Medline: 99016060. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. Zhang X, Morera S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS;

767. Kappa casein

Kappa-casein is a mammalian milk protein involved in a number of important physiological processes. In the gut, the ingested protein is split into an insoluble peptide (para kappa-casein) and a soluble hydrophilic glycopeptide (caseinomacropeptide). Caseinomacropeptide is responsible for increased efficiency of digestion, prevention of neonate hypersensitivity to ingested proteins, and inhibition of gastric pathogens. Number of members: 56

[1] Medline: 98072500. Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae. Ward TJ, Honeycutt RL, Derr JN; Genetics 1997;147:1863-1872.

768. Chitinases family 18 active site

PROSITE cross-reference(s) CHITINASE_18

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 18 (also known as classes III or V) groups a variety of proteins:

a) Chitinases from:

- Prokaryotes such as Alteromonas, Bacillus, Serratia, Streptomyces, etc.
- Plants such as Arabidopsis, cucumber, bean, tobacco, etc.

- Fungi such as Aphanocladium, Rhizopus, Saccharomyces, etc.
- Nematode (Brugia malayi).
- Insects (Manduca sexta).
- Baculoviruses (Autographa Californica Nuclear Polyhedrosis virus).

b) Other proteins:

- Hevamine, a rubber tree protein with chitinase and lysozyme activities.
- Kluyveromyces lactis killer toxin alpha subunit, which acts as a chitinase.
- Flavobacterium and Streptomyces endo-beta-N-acetylglucosaminidases (EC 3.2.1.96).
- Mammalian di-N-acetylchitobiase which is involved in the degradation of asparagine-linked glycoproteins.
- Human cartilage glycoprotein Gp-39.
- Jack bean concanavalin B (conB), a protein that has lost its catalytic activity.

Site directed mutagenesis experiments [3] and crystallographic data [4,5] have shown that a conserved glutamate is involved in the catalytic mechanism and probably acts as a proton donor. This glutamate is at the extremity of the best conserved region in these proteins.

Consensus pattern[LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E [E is the active site residue]

[1] Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992).
[2] Henrissat B. Biochem. J. 280:309-316(1991).
[3] Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H., Uchida M., Tanaka H. J. Biol. Chem. 268:18567-18572(1993).
[4] Perrakis A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S., Vorgias C.E. Structure 2:1169-1180(1994).
[5] van Scheltinga A.C.T., Kalk K.H., Beintema J.J., Dijkstra B.W. Structure 2:1181-1189(1994).

769. gag_p17. gag gene protein p17 (matrix protein).
The matrix protein forms an icosahedral shell associated with the inner membrane of the mature immunodeficiency virus. Number of members: 1598

[1] Medline: 95055757. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. Massiah MA, Starich MR, Paschall C, Summers MF, Christensen AM, Sundquist WI; J Mol Biol 1994;244:198-223.

770. GDA1/CD39 family of nucleoside phosphatases signature

PROSITE cross-reference(s); GDA1_CD39_NTPASE

A number of nucleoside diphosphate and triphosphate hydrolases as well as some yet uncharacterized proteins have been found to belong to the same family [1, 2]. This family currently consist of:

- Yeast guanosine-diphosphatase (EC 3.6.1.42) (GDPase) (gene GDA1). GDA1 is a golgi integral membrane enzyme that catalyzes the hydrolysis of GDP to GMP.
- Potato apyrase (EC 3.6.1.5) (adenosine diphosphatase) (ADPase). Apyrase acts on both ATP and ADP to produce AMP.
- Mammalian vascular ATP-diphosphohydrolase (EC 3.6.1.5) (also known as lymphoid cell activation antigen CD39).
- Toxoplasma gondii nucleoside-triphosphatases (EC 3.6.1.15) (NTPase). NTPase hydrolyses various nucleoside triphosphates to produce the corresponding nucleoside mono- and diphosphates. This enzyme is secreted into the invaded host cell into the parasitophorous vacuole, a specialized compartment where the parasite intracellulary resides.
- Pea nucleoside-triphosphatases (EC 3.6.1.15) (NTPase).

- Caenorhabditis elegans hypothetical protein C33H5.14.
- Caenorhabditis elegans hypothetical protein R07E4.4.
- Yeast chromosome V hypothetical protein YER005w.

The above uncharacterized proteins all seem to be membrane-bound.

All these proteins share a number of conserved domains. The best conserved of these domains have been selected. It is located in the central section of the proteins.

Consensus pattern[LIVM]-x-G-x(2)-E-G-x-[FY]-x-[FW]-[LIVA]-[TAG]-x-N-[HY]

[1] Handa M., Guidotti G. Biochem. Biophys. Res. Commun. 218:916-923(1996).
[2] Vasconcelos E.G., Ferreira S.T., de Carvalho T.M.U., de Souza W., Kettlun A.M., Mancilla M., Valenzuela M.A., Verjovski-Almeida S. J. Biol. Chem. 271:22139-22145(1996).

771. GTP cyclohydrolase I signatures
PROSITE cross-reference(s); GTP_CYCLOHYDROL_1_1, GTP_CYCLOHYDROL_1_2
GTP cyclohydrolase I (EC 3.5.4.16) catalyzes the biosynthesis of formic acid and dihydroneopterin triphosphate from GTP. This reaction is the first step in the biosynthesis of tetrahydrofolate in prokaryotes, of tetrahydrobiopterin in vertebrates, and of pteridine-containing pigments in insects.

GTP cyclohydrolase I is a protein of from 190 to 250 amino acid residues. The comparison of the sequence of the enzyme from bacterial and eukaryotic sources shows that the structure of this enzyme has been extremely well conserved throughout evolution [1].

Two conserved regions were selected as signature patterns. The first contains a perfectly conserved tetrapeptide which is part of the GTP-binding pocket [2], the second region also contains conserved residues involved in GTP-binding.

Consensus pattern[DEN]-[LIVM](2)-x(2)-[KRNQ]-[DEN]-[LIVM]-x(3)-[ST]-x-C-E- H-H
Consensus pattern[SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN]

[1] Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H. Biochem. Biophys. Res. Commun. 212:705-711(1995).
[2] Nar H., Huber R., Meining W., Schmid C., Weinkauf S., Bacher A. Structure 3:459-466(1995).

772. IlvC. Acetohydroxy acid isomeroreductase
Acetohydroxy acid isomeroreductase catalyses the conversion of acetohydroxy acids into dihydroxy valerates. This reaction is the second in the synthetic pathway of the essential branched side chain amino acids valine and isoleucine. Number of members: 29

[1] Medline: 97361822. The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 A resolution. Biou V, Dumas R, Cohen-Addad C, Douce R, Job D, Pebay-Peyroula E; EMBO J 1997;16:3405-3415.

773. Prokaryotic membrane lipoprotein lipid attachment site

PROSITE cross-reference(s); PROKAR_LIPOPROTEIN

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- Escherichia coli lipoprotein-34 (gene nlpB).
- Escherichia coli lipoprotein nlpC.
- Escherichia coli lipoprotein nlpD.
- Escherichia coli osmotically inducible lipoprotein B (gene osmB).
- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
- Escherichia coli copper homeostasis protein cutF (or nlpE).
- Escherichia coli plasmids traT proteins.
- Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
- Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
- Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- Fibrobacter succinogenes endoglucanase cel-3.
- Haemophilus influenzae proteins Pal and Pcp.
- Klebsiella pullulunase (gene pulA).
- Klebsiella pullulunase secretion protein pulS.
- Mycoplasma hyorhinis protein p37.

- Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC).
- Neisseria outer membrane protein H.8.
- Pseudomonas aeruginosa lipopeptide (gene lppL).
- Pseudomonas solanacearum endoglucanase egl.
- Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- Rickettsia 17 Kd antigen.
- Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
- Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
- Treponema pallidium 34 Kd antigen.
- Treponema pallidium membrane protein A (gene tmpA).
- Vibrio harveyi chitobiase (gene chb).
- Yersinia virulence plasmid protein yscJ.

- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification.

Consensus pattern{DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).

[2]Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).

[3]von Heijne G. Protein Eng. 2:531-534(1989).

[4]Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

774. Aminoacyl-transfer RNA synthetases class-II signatures

PROSITE cross-reference(s); AA_TRNA_LIGASE_II_1; AA_TRNA_LIGASE_II_2

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein

biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure.

The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].

Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns from two of these regions have been derived.

Consensus pattern[FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]
Consensus pattern[GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1]Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).
[2]Delarue M., Moras D. BioEssays 15:675-687(1993).
[3]Schimmel P. Trends Biochem. Sci. 16:1-3(1991).
[4]Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).
[5]Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).
[6]Cusack S. Biochimie 75:1077-1081(1993).
[7]Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).
[8]Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

775. X. Trans-activation protein X
This protein is found in hepadnaviruses where it is indispensable for replication. Number of members: 91

776. Thymidylate synthase active site

Thymidylate synthase (EC 2.1.1.45) [1,2] catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate. Thymidylate synthase plays an essential role in DNA synthesis and is an important target for certain chemotherapeutic drugs.

Thymidylate synthase is an enzyme of about 30 to 35 Kd in most species except in protozoan and plants where it exists as a bifunctional enzyme that includes a dihydrofolate reductase domain.

A cysteine residue is involved in the catalytic mechanism (it covalently binds the 5,6-dihydro-dUMP intermediate). The sequence around the active site of this enzyme is conserved from phages to vertebrates.

Consensus patternR-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-x-[LV] [C is the active site residue]

[1] Benkovic S.J. Annu. Rev. Biochem. 49:227-251(1980).
[2] Ross P., O'Gara F., Condon S. Appl. Environ. Microbiol. 56:2156-2163(1990).

777. Glycosyl hydrolases family 31 signatures

It has been shown [1,2,3,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Lysosomal alpha-glucosidase (EC 3.2.1.20) (acid maltase) is a vertebrate glycosidase active at low pH, which hydrolyzes alpha(1->4) and alpha(1->6) linkages in glycogen, maltose, and isomaltose.
- Alpha-glucosidase (EC 3.2.1.20) from the yeast Candida tsukunbaensis.
- Alpha-glucosidase (EC 3.2.1.20) (gene malA) from the archebacteria Sulfolobus solfataricus.
- Intestinal sucrase-isomaltase (EC 3.2.1.48 / EC 3.2.1.10) is a vertebrate membrane-bound, multifunctional enzyme complex which hydrolyzes sucrose, maltose and isomaltose. The sucrase and isomaltase domains of the enzyme are homologous (41% of amino acid identity) and have most probably evolved by duplication.
- Glucoamylase 1 (EC 3.2.1.3) (glucan 1,4-alpha-glucosidase) from various fungal species.
- Yeast hypothetical protein YBR229c.
- Fission yeast hypothetical protein SpAC30D11.01c.

An aspartic acid has been implicated [4] in the catalytic activity of sucrase, isomaltase, and lysosomal alpha-glucosidase. The region around this active residue is highly conserved and can be used as a signature pattern. A second region, which contains two conserved cysteines, has been used as an additional signature pattern.

Consensus pattern [GF]-[LIVMF]-W-x-D-M-[NSA]-E [D is the active site residue]
Consensus pattern G-[AV]-D-[LIVMTA]-C-G-[FY]-x(3)-[ST]-x(3)-L-C-x-R-W-x(2)-[LV]-[GSA]-[SA]-F-x-P-F-x-R-[DN]

[1] Henrissat B. Biochem. J. 280:309-316(1991).
[2] Kinsella B.T., Hogan S., Larkin A., Cantwell B.A. Eur. J. Biochem. 202:657-664(1991).
[3] Naim H.Y., Niermann T., Kleinhans U., Hollenberg C.P., Strasser A.W.M. FEBS Lett. 294:109-112(1991).
[4] Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A., Reuser A.J.J. J. Biol. Chem. 266:13507-13512(1991).

778. Urease signatures

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region was selected that contains two histidine that bind one of the nickel ions and the region of the active site histidine.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]
Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]

[1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).

[2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).

[3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

779. Tyrosine specific protein phosphatases signature and profiles

Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below:

Soluble PTPases.

- PTPN1 (PTP-1B).
- PTPN2 (T-cell PTPase; TC-PTP).
- PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1- like domain (see <PDOC00566>) and could act at junctions between the membrane and cytoskeleton.
- PTPN5 (STEP).
- PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup.
- PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP).
- PTPN8 (70Z-PEP).
- PTPN9 (MEG2).
- PTPN12 (PTP-G1; PTP-P19).
- Yeast PTP1.
- Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway.
- Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis.
- Fission yeast pyp3 which contributes to the dephosphorylation of cdc2.
- Yeast CDC14 which may be involved in chromosome segregation.
- Yersinia virulence plasmid PTPAses (gene yopH).
- Autographa californica nuclear polyhedrosis virus 19 Kd PTPase.

Dual specificity PTPases.

- DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185.
- DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues.
- DUSP3 (VHR).
- DUSP4 (HVH2).
- DUSP5 (HVH3).
- DUSP6 (Pyst1; MKP-3).
- DUSP7 (Pyst2; MKP-X).
- Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3.
- Yeast YVH1.
- Vaccinia virus H1 PTPase; a dual specificity phosphatase.

Receptor PTPases.

Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.

In the following table, the domain structure of known receptor PTPases is shown:

	Extracellular				Intracellular
	Ig	FN-3	CAH	MAM	PTPase
Leukocyte common antigen (LCA) (CD45)	0	2	0	0	2
Leukocyte antigen related (LAR)	3	8	0	0	2
Drosophila DLAR	3	9	0	0	2
Drosophila DPTP	2	2	0	0	2

PTP-alpha (LRP)	0	0	0	0	2
PTP-beta	0	16	0	0	1
PTP-gamma	0	1	1	0	2
PTP-delta	0	>7	0	0	2
PTP-epsilon	0	0	0	0	2
PTP-kappa	1	4	0	1	2
PTP-mu	1	4	0	1	2
PTP-zeta	0	1	1	0	2

PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.

A signature pattern was derived for PTPase domains centered on the active site cysteine.

There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.

Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue]

Notethe M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases.

Notethis documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so.

[1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991).
[2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992).
[3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991).
[4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989).
[5] Hunter T. Cell 58:1013-1016(1989).

780. Connexins signatures

Gap junctions [1] are specialized regions of the plasma membrane which consist of closely packed pairs of transmembrane channels, the connexons, through which small

molecules diffuse from a cell to a neighboring cell. Each connexon is composed of an hexamer of an integral membrane protein which is often referred to as connexin. In a given species there are a number of different, yet structurally related, tissue specific, forms of connexins. The types of connexins which are currently known are listed below.

- Connexin 56 (Cx56).
- Connexin 50 (Cx50) (lens fiber protein MP70).
- Connexin 46 (Cx46) (alpha-3).
- Connexin 45 (Cx45) (alpha-6).
- Connexin 43 (Cx43) (alpha-1).
- Connexin 40 (Cx40) (alpha-5).
- Connexin 38 (Cx38) (alpha-2).
- Connexin 37 (Cx37) (alpha-4).
- Connexin 33 (Cx33) (alpha-7).
- Connexin 32 (Cx32) (beta-1).
- Connexin 31.1 (Cx31.1) (beta-4).
- Connexin 31 (Cx31) (beta-3).
- Connexin 30.3 (Cx30.3) (beta-5).
- Connexin 26 (Cx26) (beta-2).

Structurally the connexins consist of a short cytoplasmic N-terminal domain, followed by four transmembrane segments that delimit two extracellular and one cytoplasmic loops; the C-terminal domain is cytoplasmic and its length is variable (from 20 residues in Cx26 to 260 residues in Cx56). The schematic representation of this structure is shown below.

```
NH2-***      ***       *************-COOH
    **   **  **    **
    **  **    **  **  Cytoplasmic
---**----**-----**----**----------------
    **  **    **  **  Membrane
    **  **    **  **
---**----**-----**----**----------------
    **  **    **  **  Extracellular
     ** **     ** **
      **        **
```

The sequences of the two extracellular loops are well conserved. In both loops there are three conserved cysteines which are involved in disulfide bonds. A signature patterns from each of these two loop regions has been built.

Consensus patternC-[DN]-T-x-Q-P-G-C-x(2)-V-C-[FY]-D [The three C's are involved in disulfide bonds] Consensus patternC-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P [The three C's are involved in disulfide bonds]

[1] Goodenough D.A., Goliger J.A., Paul D.L. Annu. Rev. Biochem. 65:475-502(1996).

781. Gram-positive cocci surface proteins 'anchoring' hexapeptide

Surface proteins from Gram-positive cocci contains a conserved hexapeptide located a few residues downstream of a hydrophobic C-terminal membrane anchor region which is followed by a cluster of basic amino acids [1]. This structure is represented in the following schematic representation:

```
+-----------------------------------------+-+--------+-+
| Variable length extracellular domain    |H| Anchor |B|
+-----------------------------------------+-+--------+-+
```

'H': conserved hexapeptide.

'B': cluster of basic residues.

It has been proposed that this hexapeptide sequence is responsible for a post-translational modification necessary for the proper anchoring of the proteins which bear it, to the cell wall. Proteins known to contain such hexapeptide are listed below:

- Aggregation substance from streptococcus faecalis (asa1).
- C5a peptidase from Streptococcus pyogenes (scpA).
- C protein alpha-antigen from Streptococcus agalactiae (bca).
- Cell surface antigen I/II (PAC) from Streptococcus mutans.
- Dextranase from Streptococcus downei (dex).
- Fibronectin-binding protein from Staphylococcus aureus (fnbA).
- Fimbrial subunits from Actinomyces naeslundii and viscosus.
- IgA binding protein from Streptococcus pyogenes (arp4).
- IgA binding protein (B antigen) from Streptococcus agalactiae (bag).

- IgG binding proteins from Streptococci and Staphylococcus aureus.
- Internalin A from Listeria monocytogenes (inlA).
- M proteins from streptococci.
- Muramidase-released protein from Streptococcus suis (mrp).
- Nisin leader peptide processing protease from Lactococcus lactis (nisP).
- Protein A from Staphylococcus aureus.
- Trypsin-resistant surface T protein from streptococci.
- Wall-associated protein from Streptococcus mutans (wapA).
- Wall-associated serine proteinases from Lactococcus lactis.

Consensus patternL-P-x-T-G-[STGAVDE]

[1] Schneewind O., Jones K.F., Fischetti V.A. J. Bacteriol. 172:3310-3317(1990).

782. Gamma-glutamyltranspeptidase signature

Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or water (forming glutamate). GGT plays a key role in the gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. Pseudomonas cephalosporin acylases (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region has been used as a signature pattern.

Consensus patternT-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

[1] Tate S.S., Meister A. Meth. Enzymol. 113:400-419(1985).
[2] Suzuki H., Kumagai H., Echigo T., Tochikura T. J. Bacteriol. 171:5169-5172(1989).

[3] Ishiye M., Niwa M. Biochim. Biophys. Acta 1132:233-239(1992).

783. Ferrochelatase signature

Ferrochelatase (EC 4.99.1.1) (protoheme ferro-lyase) [1,2] catalyzes the last step in heme biosynthesis: the chelation of a ferrous ion to proto-porphyrin IX, to form protoheme.

In eukaryotes, ferrochelatase is a mitochondrial protein bound to the inner membrane, whose active site faces the mitochondrial matrix. The mature form of eukaryotic ferrochelatase is composed of about 360 amino acids. In bacteria, ferrochelatase (gene hemH) [3] is a protein of from 310 to 380 amino acids.

The human autosomal dominant disease protoporphyria is due to the reduced activity of ferrochelatase.

The signature pattern for this enzyme is based on a conserved region which contains a histidine residue which could be involved in binding iron.

Consensus pattern[LIVMF](2)-x-[ST]-x-H-[GS]-[LIVM]-P-x(4,5)-[DENQKR]-x-G-[DP]-x(1,2)-Y

[1] Labbe-Bois R. J. Biol. Chem. 265:7278-7283(1990).
[2] Brenner D.A., Frasier F. Proc. Natl. Acad. Sci. U.S.A. 88:849-853(1991).
[3] Miyamoto K., Nakahigashi K., Nishimura K., Inokuchi H. J. Mol. Biol. 219:393-398(1991).

784. Cellulose-binding domain, bacterial type

The microbial degradation of cellulose and xylans requires several types of enzyme such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1].

Structurally, cellulases and xylanases generally consist of a catalytic domain joined to a cellulose-binding domain (CBD) by a short linker sequence rich in proline and/or hydroxy-amino acids.

The CBD of a number of bacterial cellulases has been shown to consist of about 105 amino acid residues [2]. Enzymes known to contain such a domain are:

- Endoglucanase (gene end1) from Butyrivibrio fibrisolvens.
- Endoglucanases A (gene cenA) and B (cenB) from Cellulomonas fimi.
- Exoglucanases A (gene cbhA) and B (cbhB) from Cellulomonas fimi.

- Endoglucanase E-2 (gene celB) from Thermomonospora fusca.
- Endoglucanase A (gene celA) from Microbispora bispora.
- Endoglucanases A (gene celA), B (celB) and C (celC) from Pseudomonas fluorescens.
- Endoglucanase A (gene celA) from Streptomyces lividans.
- Exocellobiohydrolase (gene cex) from Cellulomonas fimi.
- Xylanases A (gene xynA) and B (xynB) from Pseudomonas fluorescens.
- Arabinofuranosidase C (EC 3.2.1.55) (xylanase C) (gene xynC) from Pseudomonas fluorescens.
- Chitinase 63 (EC 3.2.1.14) from Streptomyces plicatus.
- Chitinase C from Streptomyces lividans.

The CBD domain is found either at the N-terminal or at the C-terminal extremity of these enzymes. As it is shown in the following schematic representation, there are two conserved cysteines in this CBD domain - one at each extremity of the domain - which have been shown [3] to be involved in a disulfide bond. There are also four conserved tryptophan residues which could be involved in the interaction of the CBD with polysaccharides.

```
+------------------------------------------------+
|                                    |
xCxxxxWxxxxxNxxxWxxxxxxxWxxxxxxxxWNxxxxxGxxxxxxxxxxCx
                        ********
```

'C': conserved cysteine involved in a disulfide bond. '*': position of the pattern.

Consensus patternW-N-[STAGR]-[STDN]-[LIVM]-x(2)-[GST]-x-[GST]-x(2)- [LIVMFT]-[GA]

[1] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[2] Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Protein Seq. Data Anal. 4:349-353(1991).

[3] Gilkes N.R., Claeyssens M., Aebersold R., Henrissat B., Meinke A., Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr. Eur. J. Biochem. 202:367-377(1991).

785. Amidases signature

It has been shown [1,2,3] that several enzymes from various prokaryotic and eukaryotic organisms which are involved in the hydrolysis of amides (amidases) are evolutionary related. These enzymes are listed below.

- Indoleacetamide hydrolase (EC 3.5.1.-), a bacterial plasmid-encoded enzyme that catalyzes the hydrolysis of indole-3-acetamide (IAM) into indole-3-acetate (IAA), the second step in the biosynthesis of auxins from tryptophan.
- Acetamidase from Emericella nidulans (gene amdS), an enzyme which allows acetamide to be used as a sole carbon or nitrogen source.
- Amidase (EC 3.5.1.4) from Rhodococcus sp. N-774 and Brevibacterium sp. R312 (gene amdA). This enzyme hydrolyzes propionamides efficiently, and also at a lower efficiency, acetamide, acrylamide and indoleacetamide.
- Amidase (EC 3.5.1.4) from Pseudomonas chlororaphis.
- 6-aminohexanoate-cyclic-dimer hydrolase (EC 3.5.2.12) (nylon oligomers degrading enzyme E1) (gene nylA), a bacterial plasmid encoded enzyme which catalyzes the first step in the degradation of 6-aminohexanoic acid cyclic dimer, a by-product of nylon manufacture [4].
- Glutamyl-tRNA(Gln) amidotransferase subunit A [5].
- Mammalian fatty acid amide hydrolase (gene FAAH) [6].
- A putative amidase from yeast (gene AMD2).
- Mycobacterium tuberculosis putative amidases amiA2, amiB2, amiC and amiD.

All these enzymes contain in their central section a highly conserved region rich in glycine, serine, and alanine residues. This region has been used as a signature pattern.

Consensus pattern: G-[GA]-S-[GS]-[GS]-G-x-[GSA]-[GSAVY]-x-[LIVM]-[GSA]-x(6)-[GSAT]-x-[GA]-x-[DE]-x-[GA]-x-S-[LIVM]-R-x-P-[GSAC]

[1] Mayaux J.-F., Cerbelaud E., Soubrier F., Faucher D., Petre D. J. Bacteriol. 172:6764-6773(1990).

[2] Hashimoto Y., Nishiyama M., Ikehata O., Horinouchi S., Beppu T. Biochim. Biophys. Acta 1088:225-233(1991).

[3] Chang T.-H., Abelson J. Nucleic Acids Res. 18:7180-7180(1990).

[4] Tsuchiya K., Fukuyama S., Kanzaki N., Kanagawa K., Negoro S., Okada H. J. Bacteriol. 171:3187-3191(1989).

[5] Curnow A.W., Hong K.W., Yuan R., Kim S.I., Martins O., Winkler W., Henkin T.M., Soll D. Proc. Natl. Acad. Sci. U.S.A. 94:11819-11826(1997).

[6] Cravatt B.F., Giang D.K., Mayfield S.P., Boger D.L., Lerner R.A., Gilula N.B. Nature 384:83-87(1996).

786. Glycosyl hydrolases family 10 active site

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family F [3] or as the glycosyl hydrolases family 10 [4,E1]. The enzymes which are currently known to belong to this family are listed below.

- Aspergillus awamori xylanase A (xynA).
- Bacillus sp. strain 125 xylanase (xynA).
- Bacillus stearothermophilus xylanase.
- Butyrivibrio fibrisolvens xylanases A (xynA) and B (xynB).
- Caldocellum saccharolyticum bifunctional endoglucanase/exoglucanase (celB). This protein consists of two domains; it is the N-terminal domain, which has exoglucanase activity, which belongs to this family.
- Caldocellum saccharolyticum xylanase A (xynA).
- Caldocellum saccharolyticum ORF4. This hypothetical protein is encoded in the xynABC operon and is probably a xylanase.
- Cellulomonas fimi exoglucanase/xylanase (cex).
- Clostridium stercorarium thermostable celloxylanase.
- Clostridium thermocellum xylanases Y (xynY) and Z (xynZ).
- Cryptococcus albidus xylanase.
- Penicillium chrysogenum xylanase (gene xylP).
- Pseudomonas fluorescens xylanases A (xynA) and B (xynB).
- Ruminococcus flavefaciens bifunctional xylanase XYLA (xynA). This protein consists of three domains: a N-terminal xylanase catalytic domain that belongs to family 11 of glycosyl hydrolases; a central domain composed of short repeats of Gln, Asn an Trp, and a C-terminal xylanase catalytic domain that belongs to family 10 of glycosyl hydrolases.
- Streptomyces lividans xylanase A (xlnA).

- Thermoanaerobacter saccharolyticum endoxylanase A (xynA).
- Thermoascus aurantiacus xylanase.
- Thermophilic bacterium Rt8.B4 xylanase (xynA).

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the exoglucanase from Cellulomonas fimi, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. This region has been used as a signature pattern.

Consensus pattern[GTA]-x(2)-[LIVN]-x-[IVMF]-[ST]-E-[LIY]-[DN]-[LIVMF] [E is the active site residue]

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).
[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).
[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).
[4] Henrissat B. Biochem. J. 280:309-316(1991).
[5] Tull D., Withers S.G., Gilkes N.R., Kilburn D.G., Warren R.A.J., Aebersold R. J. Biol. Chem. 266:15621-15625(1991).

787. Fructose-bisphosphate aldolase class-II signatures

Fructose-bisphosphate aldolase (EC 4.1.2.13) [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6- bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate. There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-II aldolases [2], mainly found in prokaryotes and fungi, are homodimeric enzymes which require a divalent metal ion – generally zinc - for their activity.

This family also includes the following proteins:
- Escherichia coli galactitol operon protein gatY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D- glyceraldehyde 3-phosphate.
- Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the same reaction as that of gatY.

As signature patterns for this class of enzyme, two conserved regions were selected. The first pattern is located in the first half of the sequence and contains two histidine residues that have been shown [4] to be involved in binding a zinc ion. The second is located in the C-terminal section and contains clustered acidic residues and glycines.

Consensus pattern[FYVMT]-x(1,3)-[LIVMH]-[APN]-[LIVM]-x(1,2)-[LIVM]-H-x-D-H-[GACH] [The two H's are zinc ligands]
Consensus pattern[LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E

[1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).
[2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).
[3] von der Osten C.H., Barbas C.F. III, Wong C.-H., Sinskey A.J. Mol. Microbiol. 3:1625-1637(1989).
[4] Berry A., Marshall K.E. FEBS Lett. 318:11-16(1993).

788. Prolyl oligopeptidase family serine active site

The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences.
- Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and argininyl residues.
- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.
- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.
- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).

- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus patternD-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue]

Note these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

[1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
[2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
[3] Polgar L., Szabo E.
Biol. Chem. Hoppe-Seyler 373:361-366(1992).
[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

789. Formate--tetrahydrofolate ligase signatures

Formate--tetrahydrofolate ligase (EC 6.3.4.3) (formyltetrahydrofolate synthetase) (FTHFS) is one of the enzymes participating in the transfer of one-carbon units, an essential element of various biosynthetic pathways. In many of these processes the transfers of one-carbon units are mediated by the coenzyme tetrahydrofolate (THF). Various reactions generate one-carbon derivatives of THF which can be interconverted between different oxidation states by FTHFS, methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5) and methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9).

In eukaryotes the FTHFS activity is expressed by a multifunctional enzyme, C-1-tetrahydrofolate synthase (C1-THF synthase), which also catalyzes the dehydrogenase and cyclohydrolase activities. Two forms of C1-THF synthases are known [1], one is located in the mitochondrial matrix, while the second one is cytoplasmic. In both forms the FTHFS

domain consist of about 600 amino acid residues and is located in the C-terminal section of C1-THF synthase. In prokaryotes FTHFS activity is expressed by a monofunctional homotetrameric enzyme of about 560 amino acid residues [2].

The sequence of FTHFS is highly conserved in all forms of the enzyme. As signature patterns, two regions that are almost perfectly conserved were selected. The first one is a glycine-rich segment located in the N-terminal part of FTHFS and which could be part of an ATP-binding domain [2]. The second pattern is located in the central section of FTHFS.

Consensus patternG-[LIVM]-K-G-G-A-A-G-G-G-Y
Consensus patternV-A-T-[IV]-R-A-L-K-x-[HN]-G-G

[1] Shannon K.W., Rabinowitz J.C. J. Biol. Chem. 263:7717-7725(1988).
[2] Lovell C.R., Przybyla A., Ljungdahl L.G. Biochemistry 29:5687-5694(1990).

790. Transthyretin signatures

Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis.

The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family:

- Escherichia coli hypothetical protein yedX.
- Bacillus subtilis hypothetical protein yunM.
- Caenorhabditis elegans hypothetical protein R09H10.3.
- Caenorhabditis elegans hypothetical protein ZK697.8.

Two regions were selected as signature patterns. The first located in the N-terminal extremity starts with a lysine known to be involved in binding T4. The second pattern is located in the C-terminal extremity.

Consensus pattern[KH]-[IV]-L-[DN]-x(3)-G-x-P-A-x(2)-[IV]-x-[IV] [The K binds thyroxine]
Consensus patternY-[TH]-[IV]-[AP]-x(2)-L-S-[PQ]-[FYW]-[GS]-[FY]-[QS]

[1] Schreiber G., Richardson S.J. Comp. Biochem. Physiol. 116B:137-160(1997).

791. Dihydropteroate synthase signatures

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes that are involved in the biosynthesis of folates are therefore the target of a variety of antimicrobial agents such as trimethoprim or sulfonamides.

Dihydropteroate synthase (EC 2.5.1.15) (DHPS) catalyzes the condensation of 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate to para-aminobenzoic acid to form 7,8-dihydropteroate. This is the second step in the three steps pathway leading from 6-hydroxymethyl-7,8-dihydropterin to 7,8-dihydrofolate. DHPS is the target of sulfonamides which are substrates analog that compete with para-aminobenzoic acid.

Bacterial DHPS (gene sul or folP) [1] is a protein of about 275 to 315 amino acid residues which is either chromosomally encoded or found on various antibiotic resistance plasmids. In the lower eukaryote Pneumocystis carinii, DHPS is the C-terminal domain of a multifunctional folate synthesis enzyme (gene fas) [2].

Two signature patterns for DHPS were developed, the first signature is located in the N-terminal section of these enzymes, while the second signature is located in the central section.

Consensus pattern[LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG]
Consensus pattern[GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P

[1] Slock J., Stahly D.P., Han C.-Y., Six E.W., Crawford I.P. J. Bacteriol. 172:7211-7226(1990).
[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).

792. Phosphatidylinositol 3- and 4-kinases signatures

Phosphatidylinositol 3-kinase (PI3-kinase) (EC 2.7.1.137) [1] is an enzyme that phosphorylates phosphoinositides on the 3-hydroxyl group of the inositol ring. The exact function of the three products of PI3-kinase - PI-3-P, PI-3,4-P(2) and PI-3,4,5-P(3) - is not

yet known, although it is proposed that they function as second messengers in cell signalling. Currently, three forms of PI3-kinase are known:

- The mammalian enzyme which is a heterodimer of a 110 Kd catalytic chain (p110) and an 85 Kd subunit (p85) which allows it to bind to activated tyrosine protein kinases. There are at least two different types of p100 subunits (alpha and beta).
- Yeast TOR1/DRR1 and TOR2/DRR2 [2], PI3-kinases required for cell cycle activation. Both are proteins of about 280 Kd.
- Yeast VPS34 [3], a PI3-kinase involved in vacuolar sorting and segregation. VPS34 is a protein of about 100 Kd.
- Arabidopsis thaliana and soybean VPS34 homologs.

Phosphatidylinositol 4-kinase (PI4-kinase) (EC 2.7.1.67) [4] is an enzyme that acts on phosphatidylinositol (PI) in the first committed step in the production of the second messenger inositol-1,4,5,-trisphosphate. Currently the following forms of PI4-kinases are known:

- Human PI4-kinase alpha.
- Yeast PIK1, a nuclear protein of 120 Kd.
- Yeast STT4, a protein of 214 Kd.

The PI3- and PI4-kinases share a well conserved domain at their C-terminal section; this domain seems to be distantly related to the catalytic domain of protein kinases [2]. Two signature patterns were developed from the best conserved parts of this domain.

Four additional proteins belong to this family:

- Mammalian FKBP-rapamycin associated protein (FRAP) [5], which acts as the target for the cell-cycle arrest and immunosuppressive effects of the FKBP12-rapamycin complex.
- Yeast protein ESR1 [6] which is required for cell growth, DNA repair and meiotic recombination.
- Yeast protein TEL1 which is involved in controlling telomere length.
- Yeast hypothetical protein YHR099w, a distantly related member of this family.
- Fission yeast hypothetical protein SpAC22E12.16C.

Consensus pattern[LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q

Consensus pattern[GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x-D-R-H-x(2)-N

[1] Hiles I.D., Otsu M., Volinia S., Fry M.J., Gout I., Dhand R., Panayotou G., Ruiz-Larrea F., Thompson A., Totty N.F., Hsuan J.J., Courtneidge S.A., Parker P.J., Waterfield M.D. Cell 70:419-429(1992).

[2] Kunz J., Henriquez R., Schneider U., Deuter-Reinhard M., Movva N., Hall M.N. Cell 73:585-596(1993).

[3] Schu P.V., Takegawa K., Fry M.J., Stack J.H., Waterfield M.D., Emr S.D. Science 260:88-91(1993).

[4] Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N. EMBO J. 13:2352-2361(1994).

[5] Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T., Lane W.S., Schreiber S.L. Nature 369:756-758(1994).

[6] Kato R., Ogawa H. Nucleic Acids Res. 22:3104-3112(1994).

793. FAD-dependent glycerol-3-phosphate dehydrogenase signatures

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In bacteria [1] it is associated with the utilization of glycerol coupled to respiration. In Escherichia coli, two isozymes are known: one expressed under anaerobic conditions (gene glpA) and one in aerobic conditions (gene glpD). In eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2,3].

These enzymes are proteins of about 60 to 70 Kd which contain a probable FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs from the bacterial or yeast proteins by having an EF-hand calcium-binding region (See <PDOC00018>) in its C-terminal extremity.

Two signature patterns were developed. One based on the first half of the FAD-binding domain and one which corresponds to a conserved region in the central part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

[1] Austin D., Larson T.J. J. Bacteriol. 173:101-107(1991).

[2] Roennow B., Kielland-Brandt M.C. Yeast 9:1121-1130(1993).

[3] Brown L.J., McDonald M.J., Lehn D.A., Moran S.M. J. Biol. Chem. 269:14363-14366(1994).

794. NOL1/NOP2/sun family signature

The following proteins seems to be evolutionary related:

- Mammalian proliferating-cell nucleolar antigen p120 (gene NOL1) which may play a role in the regulation of the cell cycle and the increased nucleolar activity that is associated with the cell proliferation.
- Yeast nucleolar protein NOP2 (or YNA1) which could be involved in nucleolar function during the onset of growth, and in the maintenance of nucleolar structure.
- Yeast hypothetical protein YBL024w.
- Bacterial protein sun (also known as fmu).
- Escherichia coli hypothetical protein yebU.
- Mycobacterium tuberculosis hypothetical protein MtCY21B4.24.
- Methanococcus jannaschii hypothetical protein MJ0026.

NOL1 is a protein of 855 residues, NOP2 consists of 618 residues, YBL024w of 684, sun is a protein of about 430 to 450 residues and MJ026 has 274 residues. They share a conserved central domain which contains some highly conserved regions. One of these regions was selected as a signature pattern.

Consensus pattern[FV]-D-[KRA]-[LIVMA]-L-x-D-[AV]-P-C-[ST]-[GA]

795. moaA / nifB / pqqE family signature

A number of proteins involved in the biosynthesis of metallo cofactors have been shown [1,2] to be evolutionary related. These proteins are:

- Bacterial and archebacterial protein moaA, which is involved in the biosynthesis of the molybdenum cofactor (molybdopterin; MPT).
- Arabidopsis thaliana cnx2, a protein involved in molybdopterin biosynthesis and which is highlys similar to moaA.
- Bacillus subtilis narA, which seems to be the moaA ortholog in that bacteria.

- Bacterial protein nifB (or fixZ) which is involved in the biosynthesis of the nitrogenase iron-molybdenum cofactor.
- Bacterial protein pqqE which is involved in the biosynthesis of the cofactor pyrrolo-quinoline-quinone (PQQ).
- Pyrococcus furiosus cmo, a protein involved in the synthesis of a molybdopterin-based tungsten cofactor.
- Caenorhabditis elegans hypothetical protein F49E2.1.

All these proteins share, in their N-terminal region, a conserved domain that contains three cysteines. In moaA, these cysteines have been shown [1] to be important for the biological activity. They could be inolved in the binding of an iron-sulfur cluster.

Consensus pattern[LIV]-x(3)-C-[NP]-[LIVMF]-[QRS]-C-x-[FYM]-C [The three C's are putative Fe-S ligands

[1] Menendez C., Igloi G., Henninger H., Brandsch R. Arch. Microbiol. 164:142-151(1995).
[2] Hoff T., Schnorr K.M., Meyer C., Caboche M. J. Biol. Chem. 270:6100-6107(1995).

796. Forkhead-associated (FHA) domain profile

The forkhead-associated (FHA) domain [1,E1] is a putative nuclear signalling domain found in a variety of otherwise unrelated proteins. The FHA domain comprise approximately 55 to 75 amino acids and contains three highly conserved blocks separated by divergent spacer regions. Currently it has been found in the following proteins:

- Four transcription factors that also contain a forkhead (FH) domain: mouse myocyte nuclear factor 1 (MNF1), yeast transcription factor FHL1, which probably controls pre-mRNA processing, and yeast FKH1 and FKH2. In those protein the FHA domain is located N-terminal of the DNA-binding FH domain.
- Kinase-associated protein phosphatase (KAPP) from Arabidopsis thaliana, a protein which specifically interacts with the receptor-type Ser/Thr-kinase RLK5. In KAPP, the FHA domain maps to a region that interacts with the receptor-type protein kinase RLK5 only if the kinase is phosphorylated on serine residues [2].
- Two protein kinases from yeast that are involved in mediating the nuclear response to DNA damage: DUN1 and SPK1/SAD1 [3]. The latter is the only known protein containing two copies of the FHA domain.

- Protein kinase cds1 from fission yeast contains a FHA domain and might be the ortholog of SPK1.
- Protein kinase MEK1 from yeast, which is involved in meiotic recombination.
- Human nuclear antigen Ki67 which is expressed only in proliferating cells.
- Yeast hypothetical protein YHR115c, which contains a RING-finger C-terminal of the FHA domain.
- Yeast hypothetical proteins L8083.1 and 9346.10, which contain an extensive coiled-coil region C-terminal of the FHA domain.
- Caenorhabditis elegans hypothetical protein ZK632.2.
- Caenorhabditis elegans hypothetical protein C01G6.5.
- FraH from the prokaryote Anabaena, which contains a zinc-finger motif N-terminal of the FHA domain.
- An ORF from the bacterium Streptomyces, which is on the opposite strand of the protein kinase pks1, overlapping the ORF of the kinase.

[1] Hofmann K.O., Bucher P. Trends Biochem. Sci. 20:347-349(1995).
[2] Stone J.M., Collinge M.A., Smith R.D., Horn M.A., Walker J.C. Science 266:793-795(1994).
[3] Navas T.A., Zhou Z., Elledge S.J. Cell 80:29-39(1995).

797. Ald_Xan_dh_C
Aldehyde oxidase and xanthine dehydrogenase, C terminus

[1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber R; Medline: 96072968 "Crystal structure of the xanthine oxidase-related aldehyde oxido-reductase from D. gigas." Science 1995;270:1170-1176.

Number of members: 54

798. Glyco_hydro_38
Glycosyl hydrolases family 38

Glycosyl hydrolases are key enzymes of carbohydrate metabolism.

Number of members: 20

[1] Henrissat B; Medline: 98313424; Glycosidase families" Biochem Soc Trans 1998;26:153-156.

799. HECT

HECT-domain (ubiquitin-transferase).

The name HECT comes from Homologous to the E6-AP Carboxyl Terminus.

Number of members: 43

[1] Huibregtse JM, Scheffner M, Beaudenon S, Howley PM; Medline: 95223981; A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase." Proc Natl Acad Sci U S A 1995;92:2563-2567.

800. HRDC

HRDC domain

The HRDC (Helicase and RNase D C-terminal) domain has a putative role in nucleic acid binding. Mutations in the HRDC domain cause human disease.

Number of members: 19

[1] Morozov V, Mushegian AR, Koonin EV, Bork P; Medline: 98060076; A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases" Trends Biochem Sci 1997;22:417-418.

801. Integrase

Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc binding domain. The central domain is the catalytic domain [1].The carboxyl terminal domain is a DNA binding domain [2].

Number of members: 581

[1] Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; Medline: 95099322. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases." Science 1994;266:1981-1986.
[2] Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM; Medline: 95359147; Solution structure of the DNA binding domain of HIV-1 integrase." Biochemistry 1995;34:9826-9833

802. lig_chan

Ligand-gated ion channel

This family includes the four transmembrane regions of the ionotropic glutamate receptors and NMDA receptors.

Number of members: 128

[1] Tong G, Shepherd D, Jahr CE; Medline: 95184014; Synaptic desensitization of NMDA receptors by calcineurin." Science 1995;267:1510-1512.

803. RhoGAP

RhoGAP domain

GTPase activator proteins towards Rho/Rac/Cdc42-like small GTPases.

Number of members: 97

[1] Musacchio A, Cantley LC, Harrison SC; Medline: 97121392; Crystal structure of the breakpoint cluster region-homology domain from phosphoinositide 3-kinase p85 alpha subunit." Proc Natl Acad Sci U S A 1996;93:14373-14378.
[2] Barrett T, Xiao B, Dodson EJ, Dodson G, Ludbrook SB, Nurmahomed K, Gamblin SJ, Musacchio A, Smerdon SJ, Eccleston JF; Medline: 97162209; The structure of the GTPase-activating domain from p50rhoGAP." Nature 1997;385:458-461.
[3] Rittinger K, Walker PA, Eccleston JF, Nurmahomed K, Owen D, Laue E, Gamblin SJ, Smerdon SJ; Medline: 97404320; Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP." Nature 1997;388:693-697.
[4] Boguski MS, McCormick F; Medline: 94081948; Proteins regulating Ras and its relatives." Nature 1993;366:643-654.

804. vwd

von Willebrand factor type D domain

[1] Bork P; Medline: 93327926; The modular architecture of a new family of growth regulators related to connective tissue growth factor." FEBS lett 1993;327:125-130.

Number of members: 92

805. zf-C4_Topoisom

Topoisomerase DNA binding C4 zinc finger

[1] Tse-Dinh YC, Beran-Steed RK; Medline: 89034032; Escherichia coli DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains." J Biol Chem 1988;263:15857-15859.

[2] Ahumada A, Tse-Dinh YC; Medline: 99011409; The Zn(II) binding motifs of E. coli DNA topoisomerase I is part of a high-affinity DNA binding domain." Biochem Biophys Res Commun 1998;251:509-514.

Number of members: 51

806. AIRC

AIR carboxylase

Members of this family catalyse the decarboxylation of 1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR). This family catalyse the sixth step of de novo purine biosynthesis. Some members of this family contain two copies of this domain. Number of members: 35

807. Bromodomain signature and profile

PROSITE cross-reference(s): PS00633; BROMODOMAIN_1, PS50014; BROMODOMAIN_2

The bromodomain [1,2,3] is a conserved region of about 70 amino acids found in the following proteins:

- Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1). P250 associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle.
- Human RING3, a protein of unknown function encoded in the MHC class II locus.
- Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein.
- Drosophila female sterile homeotic protein (gene fsh), required maternally for proper expression of other homeotic genes involved in pattern formation, such as Ubx.
- Drosophila brahma protein (gene brm), a protein required for the activation of multiple homeotic genes.
- Mammalian homologs of brahma. In human, three brahma-like proteins are known: SNF2a(hBRM), SNF2b, and BRG1.
- Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation
- Human peregrin (or Br140).
- Yeast BDF1 [3], a transcription factor involved in the expression of a broad class of genes including snRNAs.
- Yeast GCN5, a general transcriptional activator operating in concert with certain other DNA-binding transcriptional activators, such as GCN4, HAP2/3/4 or ADA2.
- Yeast NPS1/STH1, involved in G(2) phase control in mitosis.
- Yeast SNF2/SWI2, which is part of a complex with the SNF5, SNF6, SWI3 and ADR6/SWI1 proteins. This SWI-complex is involved in transcriptional activation.
- Yeast SPT7, a transcriptional activator of Ty elements and possibly other genes.
- Caenorhabditis elegans protein cbp-1.
- Yeast hypothetical protein YGR056w.
- Yeast hypothetical protein YKR008w.
- Yeast hypothetical protein L9638.1.

Some proteins contain a region which, while similar to some extent to a classical bromodomain, diverges from it by either lacking part of the domain or because of an insertion. These proteins are:

- Mammalian protein HRX (also known as All-1 or MLL), a protein involved in translocations leading to acute leukemias and which possibly acts as a transcriptional regulatory factor. HRX contains a region similar to the C- terminal half of the bromodomain.
- Caenorhabditis elegans hypothetical protein ZK783.4. The bromodomain of this protein has a 23 amino-acid insertion.
- Yeast protein YTA7. This protein contains a region with significant similarity to the C-terminal half of the bromodomain. As it is a member of the AAA family (see <PDOC00572>) it is also in a functionally different context.

The above proteins generally contain a single bromodomain, but some of them contain two copies, this is the case of BDF1, CCG1, fsh, RING3, YKR008w and L9638.1.

The exact function of this domain is not yet known but it is thought to be involved in protein-protein interactions and it may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

The consensus pattern that has been developed spans a major part of the bromodomain; a more sensitive detection is available through the use of a profile which spans the whole domain.

Consensus pattern[STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY]

References

[1] Haynes S.R., Doolard C., Winston F., Beck S., Trowsdale J., Dawid I.B. Nucleic Acids Res. 20:2693-2603(1992).

[2] Tamkun J.W., Deuring R., Scott M.P., Kissinger M., Pattatucci A.M., Kaufman T.C., Kennison J.A. Cell 68:561-572(1992).

[3] Tamkun J.W. Curr. Opin. Genet. Dev. 5:473-477(1995).

808. (CH) Actinin-type actin-binding domain signatures

PROSITE cross-reference(s): PS00019; ACTININ_1, PS00020; ACTININ_2

Alpha-actinin is a F-actin cross-linking protein which is thought to anchoractin to a variety of intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to membrane glycoproteins.
- In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs from the above proteins in that it contains two tandem copies of the actin-binding domain and that these copies are located in the C-terminal part of the protein.

Two conserved regions were selected as signature patterns for this type of main. The first of this region is located at the beginning of the domain, hile the second one is located in the central section and has been shown to be essential for the binding of actin.

Consensus pattern[EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N
Consensus pattern[LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x- [LIVM](2)

[1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).
[2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).
[3] Dubreuil R.R. BioEssays 13:219-226(1991).

809. (COX1) Heme-copper oxidase subunit I, copper B binding region signature
PROSITE cross-reference(s): PS00077; COX1

Heme-copper respiratory oxidases [1] are oligomeric integral membrane protein complexes that catalyze the terminal step in the respiratory chain: they transfer electrons from cytochrome c or a quinol to oxygen. Some terminal oxidases generate a transmembrane proton gradient across the plasma membrane (prokaryotes) or the mitochondrial inner membrane (eukaryotes). The enzyme

complex consists of 3-4 subunits (prokaryotes) up to 13 polypeptides (mammals) of which only the catalytic subunit (equivalent to mammalian subunit 1 (CO I)) is found in all heme-copper respiratory oxidases. The presence of a bimetallic center (formed by a high-spin heme and copper B) as well as a low-spin heme, both ligated to six conserved histidine residues near the outer side of four transmembrane spans within CO I is common to all family members [2-4].

In contrary to eukaryotes the respiratory chain of prokaryotes is branched to multiple terminal oxidases. The enzyme complexes vary in heme and copper composition, substrate type and substrate affinity. The different respiratory oxidases allow the cells to customize their respiratory systems according a variety of environmental growth conditions [1].

Recently also a component of an anaerobic respiratory chain has been found to contain the copper B binding signature of this family: nitric oxide reductase (NOR) exists in denitrifying species of Archae and Eubacteria.

Enzymes that belong to this family are:

- Mitochondrial-type cytochrome c oxidase (EC 1.9.3.1) which uses cytochrome c as electron donor. The electrons are transferred via copper A (Cu(A)) and heme a to the bimetallic center of CO I that is formed by a penta-coordinated heme a and copper B (Cu(B)). Subunit 1 contains 12 transmembrane regions. Cu(B) is said to be ligated to three of the conserved histidine residues within the transmembrane segments 6 and 7.
- Quinol oxidase from prokaryotes that transfers electrons from a quinol to the binuclear center of polypeptide I. This category of enzymes includes Escherichia coli cytochrome O terminal oxidase complex which is a component of the aerobic respiratory chain that predominates when cells are grown at high aeration.
- FixN, the catalytic subunit of a cytochrome c oxidase expressed in nitrogen-fixing bacteroids living in root nodules. The high affinity for oxygen allows oxidative phosphorylation under low oxygen concentrations. A similar enzyme has been found in other purple bacteria.

- Nitric oxide reductase (EC 1.7.99.7) from Pseudomonas stutzeri. NOR reduces nitrate to dinitrogen. It is a heterodimer of norC and the catalytic subunit norB. The latter contains the 6 invariant histidine residues and 12 transmembrane segments [5].

As a signature pattern the copper-binding region was used.

Consensus pattern[YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H [The three H's are copper B ligands]

Notecytochrome bd complexes do not belong to this family.

[1]
Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B.
J. Bacteriol. 176:5587-5600(1994).
[2]
Castresana J., Luebben M., Saraste M., Higgins D.G.
EMBO J. 13:2516-2525(1994).
[3]
Capaldi R.A., Malatesta F., Darley-Usmar V.M.
Biochim. Biophys. Acta 726:135-148(1983).
[4]
Holm L., Saraste M., Wikstrom M.
EMBO J. 6:2819-2823(1987).
[5]
Saraste M., Castresana J.
FEBS Lett. 341:1-4(1994).

810. (dehydrog_molyb) Eukaryotic molybdopterin oxidoreductases signature
PROSITE cross-reference(s): PS00559; MOLYBDOPTERIN_EUK

A number of different eukaryotic oxidoreductases that require and bind a molybdopterin cofactor have been shown [1] to share a few regions of sequence similarity. These enzymes are:

- Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of xanthine to uric acid with the concomitant reduction of NAD. Structurally, this enzyme of about 1300 amino acids consists of at least three distinct domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain (see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mo-pterin domain.
- Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its sequence and domain structure.
- Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate to nitrite. Structurally, this enzyme of about 900 amino acids consists of an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome reductase domain.
- Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to sulfate. Structurally, this enzyme of about 460 amino acids consists of an N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.

There are a few conserved regions in the sequence of the molybdopterin-binding domain of these enzymes. The pattern uses to detect these proteins is based on one of them. It contains a cysteine residue which could be involved in binding the molybdopterin cofactor.

Consensus pattern[GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-x(2)-[DE]

[1]
Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C.
Biochim. Biophys. Acta 1057:157-185(1991).

811. (DNA_ligase) ATP-dependent DNA ligase signatures
PROSITE cross-reference(s): PS00697; DNA_LIGASE_A1, PS00333; DNA_LIGASE_A2

DNA ligase (polydeoxyribonucleotide synthase) is the enzyme that joins two DNA fragments by catalyzing the formation of an internucleotide ester bond between phosphate and deoxyribose. It is active during DNA replication, DNA repair and DNA recombination. There are two forms of DNA ligase: one requires ATP (EC 6.5.1.1), the other NAD (EC 6.5.1.2).

Eukaryotic, archaebacterial, virus and phage DNA ligases are ATP-dependent. During the first step of the joining reaction, the ligase interacts with ATP to form a covalent enzyme-adenylate intermediate. A conserved lysine residue is the site of adenylation [1,2].

Apart from the active site region, the only conserved region common to all ATP-dependent DNA ligases is found [3] in the C-terminal section and contains a conserved glutamate as well as four positions with conserved basic residues.

Signature patterns were developed for both conserved regions.

Consensus pattern[EDQH]-x-K-x-[DN]-G-x-R-[GACIVM] [K is the active site residue]

Consensus patternE-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-[LIVMFY]-K
Sequences known to belong to this class detected by the patternALL, except for archebacterial DNA ligases.

[1]
Tomkinson A.E., Totty N.F., Ginsburg M., Lindahl T.
Proc. Natl. Acad. Sci. U.S.A. 88:400-404(1991).
[2]
Lindahl T., Barnes D.E.
Annu. Rev. Biochem. 61:251-281(1992).
[3]
Kletzin A.

Nucleic Acids Res. 20:5389-5396(1992).

812. (FAD_Gly3P_dh) FAD-dependent glycerol-3-phosphate dehydrogenase signatures
PROSITE cross-reference(s): PS00977; FAD_G3PDH_1, PS00978; FAD_G3PDH_2

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In bacteria [1] it is associated with the utilization of glycerol coupled to respiration. In Escherichia coli, two isozymes are known: one expressed under anaerobic conditions (gene glpA) and one in aerobic conditions (gene glpD). In eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2, 3].

These enzymes are proteins of about 60 to 70 Kd which contain a probable FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs from the bacterial or yeast proteins by having an EF-hand calcium-binding region (See <PDOC00018>) in its C-terminal extremity.

Two signature patterns were developed. One based on the first half of the FAD-binding domain and one which corresponds to a conserved region in the central part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

[1]
Austin D., Larson T.J.
J. Bacteriol. 173:101-107(1991).
[2]
Roennow B., Kielland-Brandt M.C.
Yeast 9:1121-1130(1993).
[3]
Brown L.J., McDonald M.J., Lehn D.A., Moran S.M.

J. Biol. Chem. 269:14363-14366(1994).

813. (Fapy_DNA_glyco) Formamidopyrimidine-DNA glycosylase signature
PROSITE cross-reference(s): PS01242; FPG

Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) [1] (Fapy-DNA glycosylase) (gene fpg) is a bacterial enzyme involved in DNA repair and which excise oxidized purine bases to release 2,6-diamino-4-hydroxy-5N-methylformamido-pyrimidine (Fapy) and 7,8-dihydro-8-oxoguanine (8-OxoG) residues. In addition to its glycosylase activity, FPG can also nick DNA at apurinic/apyrimidinic sites (AP sites). FPG is a monomeric protein of about 32 Kd which binds and require zinc for its activity.

The binding site for zinc seems to be located in the C-terminal part of the enzyme where fours conserved and essential [2] cysteines are located. A signature pattern was developed based on this region.

Consensus patternC-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C- x(2)-C-Q
[The four C's are putative zinc ligands]

[1]
Duwat P., de Oliveira R., Ehrlich S.D., Boiteux S.
Microbiology 141:411-417(1995).
[2]
O'Connor T.E., Graves R.J., Demurcia G., Castaing B., Laval J.
J. Biol. Chem. 268:9063-9070(1993).

814. (G_glu_transpept) Gamma-glutamyltranspeptidase signature
PROSITE cross-reference(s): PS00462; G_GLU_TRANSPEPTIDASE

Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or water (forming glutamate). GGT plays a key role in the

gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. Pseudomonas cephalosporin acylases (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region was used as a signature pattern.

Consensus patternT-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

[1]
Tate S.S., Meister A.
Meth. Enzymol. 113:400-419(1985).
[2]
Suzuki H., Kumagai H., Echigo T., Tochikura T.
J. Bacteriol. 171:5169-5172(1989).
[3]
Ishiye M., Niwa M.
Biochim. Biophys. Acta 1132:233-239(1992).

815. G-protein gamma subunit profile
PROSITE cross-reference(s): PS50058; G_PROTEIN_GAMMA

Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in

the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl-geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits.

The Caenorhabditis elegans protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain.

A profile was developed that spans the complete length of the gamma subunit.

[1]
Pennington S.R.
Protein Prof. 2:16-315(1995).

816. GNS1/SUR4 family signature
PROSITE cross-reference(s): PS01188; GNS1_SUR4

The following group of eukaryotic integral membrane proteins, whose exact function has not yet clearly been established, are evolutionary related [1]:

- Yeast GNS1 [2], a protein involved in synthesis of 1,3-beta-glucan.
- Yeast SUR4 (or APA1, SRE1) [3], a protein that could act in a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.
- Yeast hypothetical protein YJL196c.
- Caenorhabditis elegans hypothetical protein C40H1.4.
- Caenorhabditis elegans hypothetical protein D2024.3.

The proteins have from 290 to 435 amino acid residues. Structurally, they seem to be formed of three sections: a N-terminal region with two transmembrane domains, a central hydrophilic loop and a C-terminal region that contains from one to three transmembrane domains. A conserved region that contains three histidines was selected as a signature pattern. This region is located in the hydrophilic loop.

Consensus patternL-x-F-L-H-x-Y-H-H

[1]
Bairoch A.
Unpublished observations (1996).
[2]
El-Sherbeini M., Clemas J.A.
J. Bacteriol. 177:3227-3234(1995).
[3]
Garcia-Arranz M., Maldonado A.M., Mazon M.J., Portillo F.
J. Biol. Chem. 269:18076-18082(1994).

817. Immunoglobulins and major histocompatibility complex proteins signature
PROSITE cross-reference(s): PS00290; IG_MHC

The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4).

The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-

microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail.

It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. A small pattern around the C-terminal cysteine is involved in this disulfide bond which can be used to detect these category of Ig related proteins.

Consensus pattern[FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region : All, in CH2 and CH3. Ig heavy chains type Delta C region : All, in CH3. Ig heavy chains type Epsilon C region: All, in CH1, CH3 and CH4. Ig heavy chains type Gamma C region : All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region : All, in CH2, CH3 and CH4. Ig light chains type Kappa C region : In all CL except rabbit and Xenopus. Ig light chains type Lambda C region : In all CL except rabbit. MHC class I alpha chains : All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin : All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains.

[1]
Gough N.
Trends Biochem. Sci. 6:203-205(1981).
[2]
Klein J., Figueroa F.
Immunol. Today 7:41-44(1986).
[3]
Figueroa F., Klein J.
Immunol. Today 7:78-81(1986).
[4]
Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L.

Nature 282:266-270(1979).

[5]

Cushley W., Owen M.J.

Immunol. Today 4:88-92(1983).

[6]

Beck S., Barrel B.G.

Nature 331:269-272(1988).

818. (IGFBP) Insulin-like growth factor binding proteins signature

PROSITE cross-reference(s): PS00222; IGF_BINDING

The insulin-like growth factors (IGF-I and IGF-II) bind to specific binding proteins in extracellular fluids with high affinity [1,2,3]. These IGF-binding proteins (IGFBP) prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cells culture. They seem to alter the interaction of IGFs with their cell surface receptors. There are at least six different IGFBPs and they are structurally related.

The following growth-factor inducible proteins are structurally related to IGFBPs and could function as growth-factor binding proteins [4,5]:

- Mouse protein cyr61 and its probable chicken homolog, protein CEF-10.
- Human connective tissue growth factor (CTGF) and its mouse homolog, protein FISP-12.
- Vertebrate protein NOV.

As a signature pattern a conserved cysteine-rich region locatedin the N-terminal section of these proteins is used.

Consensus patternG-C-[GS]-C-C-x(2)-C-A-x(6)-C

Sequences known to belong to this class detected by the patternALL, except for IGFBP-6's.

[1]
Rechler M.M.
Vitam. Horm. 47:1-114(1993).
[2]
Shimasaki S., Ling N.
Prog. Growth Factor Res. 3:243-266(1991).
[3]
Clemmons D.R.
Trends Endocrinol. Metab. 1:412-417(1990).
[4]
Bradham D.M., Igarashi A., Potter R.L., Grotendorst G.R.
J. Cell Biol. 114:1285-1294(1991).
[5]
Maloisel V., Martinerie C., Dambrine G., Plassiart G., Brisac M., Crochet J., Perbal B.
Mol. Cell. Biol. 12:10-21(1992).

819. LMWPc : Low molecular weight phosphotyrosine protein phosphatase
Number of members: 34

[1]Medline: 94329182, The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. Su XD, Taddei N, Stefani M, Ramponi G, Nordlund P; Nature 1994;370:575-578.

820. (myosin_head) ATP/GTP-binding site motif A (P-loop)
PROSITE cross-reference(s): PS00017; ATP_GTP_A

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].

There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted is listed below:

- ATP synthase alpha and beta subunits (see <PDOC00137>).
- Myosin heavy chains.
- Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>).
- Dynamins and dynamin-like proteins (see <PDOC00362>).
- Guanylate kinase (see <PDOC00670>).
- Thymidine kinase (see <PDOC00524>).
- Thymidylate kinase (see <PDOC01034>).
- Shikimate kinase (see <PDOC00868>).
- Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>).
- ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>).
- DNA and RNA helicases [8,9,10].
- GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.).
- Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).
- Nuclear protein ran (see <PDOC00859>).
- ADP-ribosylation factors family (see <PDOC00781>).
- Bacterial dnaA protein (see <PDOC00771>).
- Bacterial recA protein (see <PDOC00131>).
- Bacterial recF protein (see <PDOC00539>).
- Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.).
- DNA mismatch repair proteins mutS family (See <PDOC00388>).
- Bacterial type II secretion system protein E (see <PDOC00567>).

Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for

adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern[AG]-x(4)-G-K-[ST]

[1]
Walker J.E., Saraste M., Runswick M.J., Gay N.J.
EMBO J. 1:945-951(1982).
[2]
Moller W., Amons R.
FEBS Lett. 186:1-7(1985).
[3]
Fry D.C., Kuby S.A., Mildvan A.S.
Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).
[4]
Dever T.E., Glynias M.J., Merrick W.C.
Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).
[5]
Saraste M., Sibbald P.R., Wittinghofer A.
Trends Biochem. Sci. 15:430-434(1990).
[6]
Koonin E.V.
J. Mol. Biol. 229:1165-1174(1993).
[7]
Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P.
J. Bioenerg. Biomembr. 22:571-592(1990).
[8]
Hodgman T.C.
Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).
[9]
Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P.
Nature 337:121-122(1989).

[10]
Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M.
Nucleic Acids Res. 17:4713-4730(1989).

821. PE: PE family

This family named after a PE motif near to the amino terminus of the domain. The PE family of proteins all contain an amino-terminal region of about 110 amino acids. The carboxyl terminus of this family are variable and fall into several classes. The largest class of PE proteins is the highly repetitive PGRS class which have a high glycine content. The function of these proteins is uncertain but it has been suggested that they may be related to antigenic variation of Mycobacterium tuberculosis [1]. Number of members: 88

[1] Medline: 98295987. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al; Nature 1998;393:537-544.

822. (RNB) Ribonuclease II family signature

PROSITE cross-reference(s): PS01175; RIBONUCLEASE_II

On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family:

- Escherichia coli and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene rnb) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction.
- Bacterial protein vacB. In Shigella flexneri, vacB has been shown to be required for the expression of virulence genes at the posttranscriptional level.
- Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase.
- Yeast protein DIS3 [2], which binds to ran (GSP1) and ehances the the

nucleotide-releasing activity of RCC1 on ran.

- Fission yeast protein dis3, which is implicated in mitotic control.
- Neurospora crassa cyt-4, a mitochondrial protein required for RNA 5' and 3' end processing and splicing.
- Yeast protein MSU1, which is involved in mitochondrial biogenesis.
- Synechocystis strain PCC 6803 protein zam [3], which control resistance to the carbonic anhydrase inhibitor acetazolamide.
- Caenorhabditis elegans hypothetical protein F48E8.6.

The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C-terminal section [4]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. A signature pattern was developed based on the core of this conserved domain.

Consensus pattern[HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ]

[1]
Zilhao R., Camelo L., Arraiano C.M.
Mol. Microbiol. 8:43-51(1993).
[2]
Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M., Nakashima N., Yanagida M., He X., Mueller U., Sazer S., Nishimoto T.
EMBO J. 15:5595-5605(1996).
[3]
Beuf L., Bedu S., Cami B., Joset F.
Plant Mol. Biol. 27:779-788(1995).
[4]
Mian I.S.
Nucleic Acids Res. 25:3187-3195(1997).

823. Src homology 2 (SH2) domain profile

PROSITE cross-reference(s): PS50001; SH2

The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6].

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets [7].

So far, SH2 domains have been identified in the following proteins:

- Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases.
- Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two copies of the SH2 domain are found in those proteins in between the catalytic 'X-' and 'Y-boxes' (see <PDOC50007>).
- Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit.
- Some vertebrate and invertebrate protein-tyrosine phosphatases.
- Mammalian Ras GTPase-activating protein (GAP).
- Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK.
- Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family.
- Miscellanous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc.
- STAT proteins (signal transducers and activators of transcription).
- Chicken tensin.
- Yeast transcriptional control protein SPT6.

The profile developed to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

[1]
Sadowski I., Stone J.C., Pawson T.
Mol. Cell. Biol. 6:4396-4408(1986).
[2]
Russel R.B., Breed J., Barton G.J.
FEBS Lett. 304:15-20(1992).
[3]
Marangere L.E.M., Pawson T.
J. Cell Sci. Suppl. 18:97-104(1994).
[4]
Pawson T., Schlessinger J.
Curr. Biol. 3:434-442(1993).
[5]
Mayer B.J., Baltimore D.
Trends Cell. Biol. 3:8-13(1993).
[6]
Pawson T.
Nature 373:573-580(1995).
[7]
Kuriyan J., Cowburn D.
Curr. Opin. Struct. Biol. 3:828-837(1993).

824. Sulfate transporters signature
PROSITE cross-reference(s): PS01130; SULFATE_TRANSP

A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

- Neurospora crassa sulfate permease II (gene cys-14).
- Yeast sulfate permeases (genes SUL1 and SUL2).
- Rat sulfate anion transporter 1 (SAT-1).
- Mammalian DTDST, a probable sulfate transporter which, in Human, is involved in the genetic disease, diastrophic dysplasia (DTD).
- Sulfate transporters 1, 2 and 3 from the legume Stylosanthes hamata.

- Human pendrin (gene PDS), which is involved in a number of hearing loss genetic diseases.
- Human protein DRA (Down-Regulated in Adenoma).
- Soybean early nodulin 70.
- Escherichia coli hypothetical protein ychM.
- Caenorhabditis elegans hypothetical protein F41D9.5.

As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.

Consensus pattern[PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]-x(3)-[GA]-[GST]-S-[KR]

[1]
Sandal N.N., Marcker K.A.
Trends Biochem. Sci. 19:19-19(1994).
[2]
Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.
Mol. Gen. Genet. 247:709-715(1995).

825. TYA: TYA transposon protein

Ty are yeast transposons. A 5.7kb transcript codes for p3 a fusion protein of TYA and TYB. The TYA protein is analogous to the gag protein of retroviruses. TYA a is cleaved to form 46kd protein which can form mature virion like particles [1]. Number of members: 59

[1] Medline: 97404699. Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like particles. Palmer KJ, Tichelaar W, Myers N, Burns NR, Butcher SJ, Kingsman AJ, Fuller SD, Saibil HR; J Virol 1997;71:6863-6868.

826. Aldolase_II

Class II Aldolase and Adducin N-terminal domain.

-!- This family includes class II aldolases and adducins which have not been ascribed any enzymatic function. Number of members: 37

References:

[1] Medline: 93294819. The spatial structure of the class II L-fuculose-1-phosphate aldolase from Escherichia coli. Dreyer MK, Schulz GE; J Mol Biol 1993;231:549-553.

[2] Medline: 96256522. Catalytic mechanism of the metal-dependent fuculose aldolase from Escherichia coli as derived from the structure. Dreyer MK, Schulz GE; J Mol Biol 1996;259:458-466.

827. CBD_2

-!- Two tryptophan residues are involved in cellulose binding.

-!- Cellulose binding domain found in bacteria. Number of members: 51

References:

[1] Medline: 95284032. Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Harvey TS; Biochemistry 1995;34:6993-7009.

828. P

A unique feature of the eukaryotic subtilisin-like proprotein convertases is the presence of an additional highly conserved sequence of approximately 150 residues (P domain) located immediately downstream of the catalytic domain.

Number of members: 91

References:

[1] Medline: 94252314. A C-terminal domain conserved in precursor processing proteases is required for intramolecular N-terminal maturation of pro-Kex2 protease. Gluschankof P, Fuller RS; EMBO J 1994;13:2280-2288.

[2] Medline: 98225190. Regulatory roles of the P domain of the subtilisin-like prohormone convertases. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF; J Biol Chem 1998;273:11107-11114.

829. Uncharacterized protein family UPF0020 signature

PROSITE cross-reference(s): PS01261; UPF0020

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Escherichia coli hypothetical protein ycbY and HI0116/15, the corresponding Haemophilus influenzae protein.
- Bacillus subtilis hypothetical protein ypsC.
- Synechocystis strain PCC 6803 hypothetical protein slr0064.
- Methanococcus jannaschii hypothetical proteins MJ0438 and MJ0710.

These are hydrophilic proteins of from 40 Kd to about 80 Kd. They can be picked up in the database by the following pattern.

Consensus patternD-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E

References:

[1] Bairoch A. Unpublished observations (1997).

830. Uncharacterized protein family UPF0031 signatures

PROSITE cross-reference(s): PS01049; UPF0031_1; PS01050; UPF0031_2

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Yeast chromosome XI hypothetical protein YKL151c.
- Caenorhabditis elegans hypothetical protein R107.2.
- Escherichia coli hypothetical protein yjeF.

- Bacillus subtilis hypothetical protein yxkO.
- Helicobacter pylori hypothetical protein HP1363.
- Mycobacterium tuberculosis hypothetical protein MtCY77.05c.
- Mycobacterium leprae hypothetical protein B229_C2_201.
- Synechocystis strain PCC 6803 hypothetical protein sll1433.
- Methanococcus jannaschii hypothetical protein MJ1586.

These are proteins of about 30 to 40 Kd whose central region is well conserved. They can be picked up in the database by the following patterns.

Consensus pattern[SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT]
Consensus pattern[GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM]

831. (ACOX)
Acyl-CoA oxidase

This is a family of Acyl-CoA oxidases EC:1.3.3.6. Acyl-coA oxidase converts acyl-CoA into trans-2-enoyl-CoA [1].

Number of members: 39

[1] Hayashi H, De Bellis L, Yamaguchi K, Kato A, Hayashi M, Nishimura M; Medline: 98192624. Molecular characterization of a glyoxysomal long chain acyl-CoA oxidase that is synthesized as a precursor of higher molecular mass in pumpkin." J Biol Chem 1998;273:8301-8307.

832. (AICARFT_IMPCHas)
AICARFT/IMPCHase bienzyme

This is a family of bifunctional enzymes catalysing the last steps in de novo purine biosynthesis. The bifunctional enzyme is found in both prokaryotes and eukaryotes. The second last step is catalysed by 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase EC:2.1.2.3 (AICARFT), this enzyme catalyses the formylation of AICAR

with 10-formyl-tetrahydrofolate to yield FAICAR and tetrahydrofolate [1]. The last step is catalysed by IMP (Inosine monophosphate) cyclohydrolase EC:3.5.4.10 (IMPCHase), cyclizing FAICAR (5-formylaminoimidazole-4-carboxamide ribonucleotide) to IMP [1].

Number of members: 22

[1] Akira T, Komatsu M, Nango R, Tomooka A, Konaka K, Yamauchi M, Kitamura Y, Nomura S, Tsukamoto I; Medline: 97473523 Molecular cloning and expression of a rat cDNA encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase" [published erratum appears in Gene 1998 Feb 27;208(2):337] Gene 1997;197:289-293.

[2] Rayl EA, Moroson BA, Beardsley GP; Medline: 96147205 The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping." J Biol Chem 1996;271:2225-2233.

833. (AOX)

Alternative oxidase

The alternative oxidase is used as a second terminal oxidase in the mitochondria, electrons are transfered directly from reduced ubiquinol to oxygen forming water [2]. This is not coupled to ATP synthesis and is not inhibited by cyanide, this pathway is a single step process [1]. In rice the transcript levels of the alternative oxidase are increased by low temperature [1].

Number of members: 27

[1] Ito Y, Saisho D, Nakazono M, Tsutsumi N, Hirai A; Medline: 98086211 Transcript levels of tandem-arranged alternative oxidase genes in rice are increased by low temperature." Gene 1997;203:121-129.

[2] Li Q, Ritzel RG, McLean LL, McIntosh L, Ko T, Bertrand H, Nargang FE; Medline: 96366413 Cloning and analysis of the alternative oxidase gene of Neurospora crassa." Genetics 1996;142:129-140.

834. (APH)

Protein kinases signatures and profile

Cross-reference(s): PS00107; PROTEIN_KINASE_ATP, PS00108; PROTEIN_KINASE_ST, PS00109; PROTEIN_KINASE_TYR, PS50011; PROTEIN_KINASE_DOM

Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of these regions have been selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; two signature patterns were derived for that region: one specific for serine/ threonine kinases and the other for tyrosine kinases. A profile was developed which is based on the alignment in [1] and covers the entire catalytic domain.

Consensus pattern: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]

Sequences known to belong to this class detected by the pattern the majority of known protein kinases but it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]

Sequences known to belong to this class detected by the pattern. Most serine/ threonine specific protein kinases with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3) [D is an active site residue] tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses ganciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. Sequences known to belong to this class detected by the profile ALL, except for three viral kinases. This profile also detects receptor guanylate cyclases (see <PDOC00430>) and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects Arabidopsis thaliana kinase- like protein TMKL1 which seems to have lost its catalytic activity.

Note if a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Note eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus [11] and Yersinia pseudotuberculosis. Note the patterns shown above has been updated since their publication in [7]. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

References

[1] Hanks S.K., Hunter T., FASEB J. 9:576-596(1995).

[2] Hunter T., Meth. Enzymol. 200:3-37(1991).

[3] Hanks S.K., Quinn A.M., Meth. Enzymol. 200:38-62(1991).

[4] Hanks S.K., Curr. Opin. Struct. Biol. 1:369-383(1991).

[5] Hanks S.K., Quinn A.M., Hunter T., Science 241:42-52(1988).

[6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor, S.S., Sowadski J.M., Science 253:407-414(1991).

[7] Bairoch A., Claverie J.-M., Nature 331:22(1988).

[8] Benner S., Nature 329:21-21(1987).

[9] Kirby R., J. Mol. Evol. 30:489-492(1992).

[10] Littler E., Stuart A.D., Chee M.S., Nature 358:160-162(1992).

[11] Munoz-Dorado J., Inouye S., Inouye M., Cell 67:995-1006(1991).

835. (Asp_Glu_race)

Aspartate and glutamate racemases signatures

Cross-reference(s) PS00923; ASP_GLU_RACEMASE_1 PS00924; ASP_GLU_RACEMASE_2

Aspartate racemase (EC 5.1.1.13) and glutamate racemase (EC 5.1.1.3) are two evolutionary related bacterial enzymes that do not seem to require a cofactor for their activity [1]. Glutamate racemase, which interconverts L-glutamate into D-glutamate, is required for the biosynthesis of peptidoglycan and some peptide-based antibiotics such as gramicidin S. In addition to characterized aspartate and glutamate racemases, this family also includes a hypothetical protein from Erwinia carotovora and one from Escherichia coli (ygeA). Two conserved cysteines are present in the sequence of these enzymes. They are expected to play a role in catalytic activity by acting as bases in proton abstraction from the substrate. Signature patterns were developed for both cysteines.

Consensus pattern: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK]

Consensus pattern: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM]

[1] Gallo K.A., Knowles J.R., Biochemistry 32:3981-3990(1993).

836. (ATP-sulfurylase)

ATP-sulfurylase

This family consists of ATP-sulfurylase or sulfate adenylyltransferase EC:2.7.7.4 some of which are part of a bifunctional polypeptide chain associated with adenosyl phosphosulphate (APS) kinase APS_kinase. Both enzymes are required for PAPS (phosphoadenosine-phosphosulfate) synthesis from inorganic sulphate [2]. ATP sulfurylase catalyses the synthesis of adenosine-phosphosulfate APS from ATP and inorganic sulphate [1].

Number of members: 37

[1] Kurima K, Warman ML, Krishnan S, Domowicz M, Krueger RC Jr, Deyrup A, Schwartz NB; Medline: 98337975 A member of a family of sulfate-activating enzymes causes murine brachymorphism" [published erratum appears in Proc Natl Acad Sci U S A 1998 Sep 29;95(20):12071] Proc Natl Acad Sci U S A 1998;95:8681-8685.

[2] Rosenthal E, Leustek T; Medline: 96096529 A multifunctional Urechis caupo protein, PAPS synthetase, has both ATP sulfurylase and APS kinase activities." Gene 1995;165:243-248.

837. (ATP-synt_F)

ATP synthase (F/14-kDa) subunit

This family includes 14-kDa subunit from vATPases [1], which is in the peripheral catalytic part of the complex [2]. The family also includes archaebacterial ATP synthase subunit F [3].

Number of members: 23

[1] Guo Y, Kaiser K, Wieczorek H, Dow JA; Medline: 96269411 The Drosophila melanogaster gene vha14 encoding a 14-kDa F-subunit of the vacuolar ATPase." Gene 1996;172:239-243.

[2] Peng SB, Crider BP, Tsai SJ, Xie XS, Stone DK; Medline: 96216416 Identification of a 14-kDa subunit associated with the catalytic sector of clathrin-coated vesicle H+-ATPase." J Biol Chem 1996;271:3324-3327.

[3] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

838. (CBD_4)
Starch binding domain

Number of members: 48

839. (CbiX)

The function of CbiX is uncertain, however it is found in cobalamin biosynthesis operons and so may have a related function. Some CbiX proteins contain a striking histidine-rich region at their C-terminus, which suggests that it might be involved in metal chelation [1].

Number of members: 6

[1] Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Medline: 98416126 Cobalamin (vitamin B12) biosynthesis: identification and characterization of a Bacillus megaterium cobI operon." Biochem J 1998;335:159-166.

840. (Complex1_51K)

Respiratory-chain NADH dehydrogenase 51 Kd subunit signatures Cross-reference(s) PS00644; COMPLEX1_51K_1 PS00645; COMPLEX1_51K_2

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 51 Kd (in mammals),

which is the second largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind to NAD, FMN, and a 2Fe-2S cluster.

The 51 Kd subunit is highly similar to [3,4]:

- Subunit alpha of Alcaligenes eutrophus NAD-reducing hydrogenase (gene hoxF) which also binds to NAD, FMN, and a 2Fe-2S cluster.
- Subunit NQO1 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
- Subunit F of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoF).

The 51 Kd subunit and the bacterial hydrogenase alpha subunit contains three regions of sequence similarities. The first one most probably corresponds to the NAD-binding site, the second to the FMN-binding site, and the third one, which contains three cysteines, to the iron-sulfur binding region. Signature patterns have been developed for the FMN-binding and for the 2Fe-2S binding regions.

Consensus pattern: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]- S
Consensus pattern: E-S-C-G-x-C-x-P-C-R-x-G [The three C's are putative 2Fe-2S ligands]

[1] Ragan C.I., Curr. Top. Bioenerg. 15:1-36(1987).
[2] Weiss H., Friedrich T., Hofhaus G., Preis D., Eur. J. Biochem. 197:563-576(1991).
[3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).
[4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H., J. Mol. Biol. 233:109-122(1993).

841. (DAP_epimerase)
Diaminopimelate epimerase signature

Cross-reference(s) PS01326; DAP_EPIMERASE
Diaminopimelate epimerase (EC 5.1.1.7) catalyzes the isomeriazation of L,L- to D,L-meso-diaminopimelate in the biosynthetic pathway leading from aspartate to lysine. This enzyme is a protein of about 30 Kd. Two conserved cysteines seem [1] to function as the acid and base in the catalytic mechanism. As a signature pattern, the region surrounding the first of these two active site cysteines were selected.

Consensus pattern: N-x-D-G-S-x(4)-C-G-N-[GA]-x-R [C is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for an Anabaena dapF which has a Ser instead of the active site Cys.

[1] Cirilli M., Zheng R., Scapin G., Blanchard J.S., Biochemistry 37:16452-16458(1998).

842. (DNA_gyraseB_C)
DNA topoisomerase II signature

Cross-reference(s) PS00177; TOPOISOMERASE_II
DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

```
<----------------About-1400-residues---------------------->

[----------Protein 39-*-----][----Protein 52----]         Phage T4
[----------gyrB-------*-----][--------gyrA----------------]  Prokaryote II
                                                    Archaebacteria
[----------parE-------*-----][--------parD----------------]  Prokaryote IV
[-------------------*---------------------------------]   Eukaryote and
```

'*': Position of the pattern.

As a signature pattern for this family of proteins, a region that contains a highly conserved pentapeptide was selected. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

Consensus pattern: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG]

[1] Sternglanz R., Curr. Opin. Cell Biol. 1:533-535(1990).
[2] Bjornsti M.-A., Curr. Opin. Struct. Biol. 1:99-103(1991).
[3] Sharma A., Mondragon A., Curr. Opin. Struct. Biol. 5:39-47(1995).
[4] Roca J., Trends Biochem. Sci. 20:156-160(1995).

843. (DUF16)

Protein of unknown function

The function of this protein is unknown. It appears to only occur in Mycoplasma pneumoniae.

Number of members: 26

[1] Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R; Medline: 97105885 Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae." Nucleic Acids Res 1996;24:4420-4449.

844. (DUF21)

Domain of unknown function

This transmembrane region has no known function. Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not

contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

845. (DUF56)

Integral membrane protein

The members of this family are putative integral membrane proteins. The function of the family is unknown, however the family includes Sec59 from yeast. Sec59 is a dolichol kinase EC:2.7.1.108, but it is not clear if the enzymatic activity resides in this region or its N terminal region.

Number of members: 13

846. (DUF94)

Domain of unknown function

The function of this domain is unknown. It is found in both eukaryotes and archaebacteria. The alignment contains a completely conserved aspartate residue that may be functionally important. The eukaryotic domains contains three conserved cysteines and a histidine that might be metal binding, however these are absent in the archaebacterial proteins.

Number of members: 9

847. (FF)

FF domain

This domain may be involved in protein-protein interaction [1].

Number of members: 42

[1] Bedford MT, Leder P; Medline: 99322199 The FF domain: a novel motif that often accompanies WW domains." Trends Biochem Sci 1999;24:264-265.

848. (FLO_LFY)

Floricaula / Leafy protein

This family consists of various plant development proteins which are homologues of floricaula (FLO) and Leafy (LFY) proteins which are floral meristem identity proteins. Mutations in the sequences of these proteins affect flower and leaf development.

Number of members: 16

[1] Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N; Medline: 97411151 UNIFOLIATA regulates leaf and flower morphogenesis in pea." Curr Biol 1997;7:581-587.

[2] Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM; Medline: 92274452 LEAFY controls floral meristem identity in Arabidopsis." Cell 1992;69:843-859.

849. (G-patch)

G-patch domain

This domain is found in a number of RNA binding proteins, and is also found in proteins that contain RNA binding domains. This suggests that this domain may have an RNA binding function. This domain has seven highly conserved glycines.

Number of members: 47

[1] Aravind L, Koonin EV; Medline: 10470032 G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins." Trends Biochem Sci 1999;24:342-344.

850. (Gram-ve_porins)
General diffusion Gram-negative porins signature

Cross-reference(s) PS00576; GRAM_NEG_PORIN

The outer membrane of Gram-negative bacteria acts as a molecular filter for hydrophilic compounds. Proteins, known as porins [1], are responsible for the 'molecular sieve' properties of the outer membrane. Porins form large water- filled channels which allows the diffusion of hydrophilic molecules into the periplasmic space. Some porins form general diffusion channels that allows any solutes up to a certain size (that size is known as the exclusion limit) to cross the membrane, while other porins are specific for a solute and contain a binding site for that solute inside the pores (these are known as selective porins). As porins are the major outer membrane proteins, they also serve as receptor sites for the binding of phages and bacteriocins. General diffusion porins generally assemble as trimer in the membrane and the transmembrane core of these proteins is composed exclusively of beta strands [2]. It has been shown [3] that a number of general porins are evolutionary related, these porins are:

- Enterobacteria phoE.
- Enterobacteria ompC.
- Enterobacteria ompF.
- Enterobacteria nmpC.
- Bacteriophage PA-2 LC.
- Neisseria PI.A.
- Neisseria PI.B.

As a signature pattern a conserved region was selected, located in the C-terminal part of these proteins, which spans two putative transmembrane beta strands.

Consensus pattern: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]- V

[1] Benz R., Bauer K., Eur. J. Biochem. 176:1-19(1988).

[2] Jap B.K., Walian P.J., Q. Rev. Biophys. 23:367-403(1990).

[3] Jeanteur D., Lakey J.H., Pattus F., Mol. Microbiol. 5:2153-2164(1991).

851. (HlyD)

HlyD family secretion proteins signature

Cross-reference(s) PS00543; HLYD_FAMILY

Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, require the help of two or more proteins for their secretion across the cell envelope. Amongst which a protein belonging to the ABC transporters family (see the relevant entry <PDOC00185>) and a protein belonging to a family which is currently composed [1 to 5] of the following members:

Gene	Species	Protein which is exported
hlyD	Escherichia coli	Hemolysin
appD	A.pleuropneumoniae	Hemolysin
lcnD	Lactococcus lactis	Lactococcin A
lktD	A.actinomycetemcomitans Pasteurella haemolytica	Leukotoxin
rtxD	A.pleuropneumoniae	Toxin-III
cyaD	Bordetella pertussis	Calmodulin-sensitive adenylate cyclase-hemolysin (cyclolysin)
cvaA	Escherichia coli	Colicin V
prtE	Erwinia chrysanthemi	Extracellular proteases B and C
aprE	Pseudomonas aeruginosa	Alkaline protease
emrA	Escherichia coli	Drugs and toxins
yjcR	Escherichia coli	Unknown

These proteins are evolutionary related and consist of from 390 to 480 amino acid residues. They seem to be anchored in the inner membrane by a N-terminal transmembrane region. Their exact role in the secretion process is not yet known. The C-terminal section of these proteins is the best conserved region; a signature pattern from that region was derived.

Consensus pattern: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-[LIVMFYW](2)-x-[LIVMFYW](3)

Sequences known to belong to this class detected by the pattern ALL, except for emrA and yjcR.

References:

[1] Gilson L., Mahanty H.K., Kolter R., EMBO J. 9:3875-3884(1990).

[2] Letoffe S., Delepelaire P., Wandersman C., EMBO J. 9:1375-1382(1990).

[3] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L., Appl. Environ. Microbiol. 58:1952-1961(1992).

[4] Duong F., Lazdunski A., Cami B., Murgier M., Gene 121:47-54(1992).

[5] Lewis K., Trends Biochem. Sci. 19:119-123(1994).

852. (IBR)

In Between Ring fingers

The IBR (In Between Ring fingers) domain is found to occur between pairs of ring fingers (zf-C3HC4). The function of this domain is unknown. This domain has also been called the C6HC domain and DRIL (for double RING finger linked) domain [2].

Number of members: 25

[1] Morett E, Bork P; Medline: 10366851 A novel transactivation domain in parkin."Trends Biochem Sci 1999;24:229-231.

[2] van der Reijden BA, Erpelinck-Verschueren CA, Lowenberg B, Jansen JH; Medline: 99349709 TRIADs: a new class of proteins with a novel cysteine-rich signature." Protein Sci 1999;8:1557-1561.

853. (IPPT)

IPP transferase

[1] Durand JM, Bjork GR, Kuwae A, Yoshikawa M, Sasakawa C; Medline: 97440126 The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of Shigella flexneri is required for expression of virulence genes." J Bacteriol 1997;179:5777-5782.
[2] Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC, Hopper AK; Medline: 94187700 Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol." Mol Cell Biol 1994;14:2298-2306.
[3] Gillman EC, Slusher LB, Martin NC, Hopper AK; Medline: 91203856 MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA." Mol Cell Biol 1991;11:2382-2390.

854. (KE2)
KE2 family protein

The function of members of this family is unknown, although they have been suggested to contain a DNA binding leucine zipper motif [2].

Number of members: 9

[1] Ha H, Abe K, Artzt K; Medline: 92084131 Primary structure of the embryo-expressed gene KE2 from the mouse H-2K region." Gene 1991;107:345-346.
[2] Shang HS, Wong SM, Tan HM, Wu M; Medline: 95129859 YKE2, a yeast nuclear gene encoding a protein showing homology to mouse KE2 and containing a putative leucine-zipper motif." Gene 1994;151:197-201.

855. (Lipoprotein_6)
Prokaryotic membrane lipoprotein lipid attachment site

Cross-reference(s) PS00013; PROKAR_LIPOPROTEIN
In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which

a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- Escherichia coli lipoprotein-34 (gene nlpB).
- Escherichia coli lipoprotein nlpC.
- Escherichia coli lipoprotein nlpD.
- Escherichia coli osmotically inducible lipoprotein B (gene osmB).
- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
- Escherichia coli copper homeostasis protein cutF (or nlpE).
- Escherichia coli plasmids traT proteins.
- Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
- Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
- Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- Fibrobacter succinogenes endoglucanase cel-3.
- Haemophilus influenzae proteins Pal and Pcp.
- Klebsiella pullulunase (gene pulA).
- Klebsiella pullulunase secretion protein pulS.
- Mycoplasma hyorhinis protein p37.
- Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC).
- Neisseria outer membrane protein H.8.
- Pseudomonas aeruginosa lipopeptide (gene lppL).
- Pseudomonas solanacearum endoglucanase egl.
- Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- Rickettsia 17 Kd antigen.
- Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
- Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
- Treponema pallidium 34 Kd antigen.
- Treponema pallidium membrane protein A (gene tmpA).

- Vibrio harveyi chitobiase (gene chb).
- Yersinia virulence plasmid protein yscJ.
- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper-binding protein. This is the first archaebacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification were derived.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1)

The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

References

[1] Hayashi S., Wu H.C., J. Bioenerg. Biomembr. 22:451-471(1990).
[2] Klein P., Somorjai R.L., Lau P.C.K., Protein Eng. 2:15-20(1988).
[3] von Heijne G., Protein Eng. 2:531-534(1989).
[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

856. (Lipoprotein_7)
Adhesin lipoprotein

This family consists of the p50 and variable adherence-associated antigen (Vaa) adhesins from Mycoplasma hominis. M. hominis is a mycoplasma associated with human urogenital diseases, pneumonia, and septic arthritis [1]. An adhesin is a cell surface molecule that mediates adhesion to other cells or to the surrounding surface or substrate. The Vaa antigen is a 50-kDa surface lipoprotein that has four tandem repetitive DNA sequences encoding a periodic peptide structure, and is highly immunogenic in the human host [1]. p50 is also a 50-

kDa lipoprotein, having three repeats A,B and C, that may be a tetramer of 191-kDa in its native environment [2].

Number of members: 18

[1] Zhang Q, Wise KS; Medline: 96294788 Molecular basis of size and antigenic variation of a Mycoplasma hominis adhesin encoded by divergent vaa genes. Infect Immun 1996;64:2737-2744.

[2] Henrich B, Kitzerow A, Feldmann RC, Schaal H, Hadding U; Medline: 97047675 Repetitive elements of the Mycoplasma hominis adhesin p50 can be differentiated by monoclonal antibodies." Infect Immun 1996;64:4027-4034.

857. (MaoC_like)

MaoC like domain

The MaoC protein is found to share similarity with a wide variety of enzymes; estradiol 17 beta-dehydrogenase 4, peroxisomal hydratase-dehydrogenase-epimerase, fatty acid synthase beta subunit. All these enzymes contain other domains. This domain is also present in the NodN nodulation protein N. No specific function has been assigned to this region of any of these proteins. The maoC gene is part of a operon with maoA which is involved in the synthesis of monoamine oxidase [1].

Number of members: 46

[1] Sugino H, Sasaki M, Azakami H, Yamashita M, Murooka Y Medline: 96235221 A monoamine-regulated Klebsiella aerogenes operon containing the monoamine oxidase structural gene (maoA) and the maoC gene." J Bacteriol 1992;174:2485-2492.

858. (MSP)

Manganese-stabilizing protein / photosystem II polypeptide

This family consists of the 33 KDa photosystem II polypeptide from the oxygen evolving complex (OEC) of plants and cyanobacteria. The protein is also known as the manganese-stabilizing protein as it is associated with the manganese complex of the OEC and may provide the ligands for the complex [1].

Number of members: 17

[1] Philbrick JB, Zilinskas BA; Medline: 88334494 "Cloning, nucleotide sequence and mutational analysis of the gene encoding the Photosystem II manganese-stabilizing polypeptide of Synechocystis 6803." Mol Gen Genet 1988;212:418-425.

859. (NAC)

[1] Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Medline: 99342100 Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell." Genome Res 1999;9:608-628.

Number of members: 27

860. (Nop)
Putative snoRNA binding domain

This family consists of various Pre RNA processing ribonucleoproteins. The function of the aligned region is unknown however it may be a common RNA or snoRNA or Nop1p binding domain. Nop5p (Nop58p) Swiss:Q12499 from yeast is the protein component of a ribonucleoprotein protein required for pre-18s rRNA processing and is suggested to function with Nop1p in a snoRNA complex [1]. Nop56p Swiss:O00567 and Nop5p interact with Nop1p and are required for ribosome biogenesis [2]. Prp31p Swiss:p49704 is required for pre-mRNA splicing in S. cerevisiae [3].

Number of members: 23

[1] Wu P, Brockenbrough JS, Metcalfe AC, Chen S, Aris JP; Medline: 98298165 Nop5p is a small nucleolar ribonucleoprotein component required for pre- 18 S rRNA processing in yeast." J Biol Chem 1998;273:16453-16463.

[2] Gautier T, Berges T, Tollervey D, Hurt E;Medline: 8038777 Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis." Mol Cell Biol 1997;17:7088-7098.

[3] Weidenhammer EM, Singh M, Ruiz-Noriega M, Woolford JL Jr; Medline: 96184869 The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae." Nucleic Acids Res 1996;24:1164-1170.

861. (Nramp)

Natural resistance-associated macrophage protein

The natural resistance-associated macrophage protein (NRAMP) family consists of Nramp1, Nramp2, and yeast proteins Smf1 and Smf2. The NRAMP family is a novel family of functional related proteins defined by a conserved hydrophobic core of ten transmembrane domains [5]. This family of membrane proteins are divalent cation transporters. Nramp1 is an integral membrane protein expressed exclusively in cells of the immune system and is recruited to the membrane of a phagosome upon phagocytosis [1]. By controlling divalent cation concentrations Nramp1 may regulate the interphagosomal replication of bacteria [1]. Mutations in Nramp1 may genetically predispose an individual to susceptibility to diseases including leprosy and tuberculosis conversely this might however provide protection form rheumatoid arthritis [1]. Nramp2 is a multiple divalent cation transporter for Fe2+, Mn2+ and Zn2+ amongst others it is expressed at high levels in the intestine; and is major transferrin-independent iron uptake system in mammals [1]. The yeast proteins Smf1 and Smf2 may also transport divalent cations [3].

Number of members: 36

[1] Govoni G, Gros P; Medline: 98383996 Macrophage NRAMP1 and its role in resistance to microbial infections." Inflamm Res 1998;47:277-284.

[2] Agranoff DD, Krishna S Medline: 98294035 Metal ion homeostasis and intracellular parasitism." Mol Microbiol 1998;28:403-412.

[3] Pinner E, Gruenheid S, Raymond M, Gros P; Medline: 98030569 Functional complementation of the yeast divalent cation transporter family SMF by NRAMP2, a member of the mammalian natural resistance- associated macrophage protein family." J Biol Chem 1997;272:28933-28938.

[4] Cellier M, Belouchi A, Gros P; Medline: 96402487 Resistance to intracellular infections: comparative genomic analysis of Nramp." Trends Genet 1996;12:201-204.

[5] Cellier M, Prive G, Belouchi A, Kwan T, Rodrigues V, Chia W, Gros P; Medline: 96036029 Nramp defines a family of membrane proteins." Proc Natl Acad Sci U S A 1995;92:10089-10093.

862. (NTP_transf_2)

Nucleotidyltransferase domain

Members of this family belong to a large family of nucleotidyltransferases [1].

Number of members: 83

[1] Holm L, Sander C; Medline: 96005605 DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily." Trends Biochem Sci 1995;20:345-347.

863. (Paramyxo_P)

Paramyxovirus P phosphoprotein

This family consists of paramyxovirus P phosphoprotein from sendai virus and human and bovine parainfluenza viruses. The P protein is an essential part of the viral RNA polymerase complex formed form the P and L proteins [1]. The exact role of the P protein in this complex in unknown but it is involved in multiple protein-protein interactions and binding the polymerase complex to the nucleocapsid or ribonucleoprotein template [1]. It also appears to be important for the proper folding of the L protein [1]. The paramyxoviruses have a negative sense ssRNA genome [1].

Number of members: 15

[1] Bowman MC, Smallwood S, Moyer SA; Medline: 99329169 Dissection of Individual Functions of the Sendai Virus Phosphoprotein in Transcription." J Virol 1999;73:6474-6483.
[2] Matsuoka Y, Curran J, Pelet T, Kolakofsky D, Ray R, Compans RW; Medline: 91237868 The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine-rich V protein." J Virol 1991;65:3406-3410.

864. (Patatin)

This family consists of various patatin glycoproteins from plants. The patatin protein accounts for up to 40% of the total soluble protein in potato tubers [2]. Patatin is a storage protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids [2].

Number of members: 21

[1] Banfalvi Z, Kostyal Z, Barta E; Medline: 95107249 Solanum brevidens possesses a non-sucrose-inducible patatin gene." Mol Gen Genet 1994;245:517-522.
[2] Mignery GA, Pikaard CS, Park WD; Medline: 88226014 Molecular characterization of the patatin multigene family of potato." Gene 1988;62:27-44.

865. (Pentapeptide_2)
Pentapeptide repeats (8 copies)

These repeats are found in many mycobacterial proteins. These repeats are most common in the PPE family of proteins, where they are found in the MPTR subfamily of PPE proteins. The function of these repeats is unknown. The repeat can be approximately described as XNXGX, where X can be any amino acid. These repeats are similar to Pentapeptide [1], however it is not clear if these two families are structurally related.

Number of members: 362

[1] Bateman A, Murzin A, Teichmann SA; Medline: 98318059 Structure and distribution of pentapeptide repeats in bacteria." Protein Sci 1998;7:1477-1480.
[2] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG; Medline: 98295987 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." Nature 1998;393:537-544.

866. (Peptidase_C13)
Peptidase C13 family

This family of peptidases is known as the hemoglobinase family because it contains a globin degrading enzyme from blood parasites Swiss:P42665. However relatives are found in plants and other organisms that have other functions. Members of this family are asparaginyl peptidases [1].

Number of members: 26

[1] Chen JM, Dando PM, Rawlings ND, Brown MA, Young NE, Stevens RA, Hewitt E, Watts C, Barrett AJ; Medline: 97218252 Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase." J Biol Chem 1997;272:8090-8098.

867. (Pro_dh)
Proline dehydrogenase

Number of members: 25

[1] Ling M, Allen SW, Wood JM; Medline: 95055736 Sequence analysis identifies the proline dehydrogenase and delta 1- pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli PutA protein." J Mol Biol 1994;243:950-956.

868. (PsbP)

This family consists of the 23 kDa subunit of oxygen evolving system of photosystem II or PsbP from various plants (where it is encoded by the nuclear genome) and Cyanobacteria. The 23 KDa PsbP protein is required for PSII to be fully operational in vivo, it increases the affinity of the water oxidation site for Cl- and provides the conditions required for high affinity binding of Ca2+ [2].

Number of members: 25

[1] Rova EM, Mc Ewen B, Fredriksson PO, Styring S; Medline: 97067138 Photoactivation and photoinhibition are competing in a mutant of Chlamydomonas reinhardtii lacking the 23-kDa extrinsic subunit of photosystem II." J Biol Chem 1996;271:28918-28924.
[2] Kochhar A, Khurana JP, Tyagi AK; Medline: 97191538 Nucleotide sequence of the psbP gene encoding precursor of 23-kDa polypeptide of oxygen-evolving complex in Arabidopsis thaliana and its expression in the wild-type and a constitutively photomorphogenic mutant." DNA Res 1996;3:277-285.

869. (PUA)

The PUA domain named after PseudoUridine synthase and Archaeosine transglycosylase, was detected in archaeal and eukaryotic pseudouridine synthases, archaeal archaeosine synthases, a family of predicted ATPases that may be involved in RNA modification, a family of predicted archaeal and bacterial rRNA methylases. Additionally, the PUA domain was detected in a family of eukaryotic proteins that also contain a domain homologous to the translation initiation factor eIF1/SUI1; these proteins may comprise a novel type of translation factors. Unexpectedly, the PUA domain was detected also in bacterial and yeast glutamate kinases; this is compatible with the demonstrated role of these enzymes in the

regulation of the expression of other genes [1]. It is predicted that the PUA domain is an RNA binding domain.

Number of members: 48

[1] Aravind L, Koonin EV; Medline: 99193178 Novel predicted RNA-binding domains associated with the translation machinery." J Mol Evol 1999;48:291-302.

870. (RF1)
eRF1-like proteins

Members of this family are peptide chain release factors. The eukaryotic Release Factor 1 proteins (eRF1s) are involved in termination of translation. The eRF1 protein is functional for all stop codons and appears to abolish read-through of these codons. This family also includes other proteins for which the precise molecular function is unknown. Many of them are from Archaebacteria. These proteins may also be involved in translation termination but this awaits experimental verification. Number of members: 25

[1] Frolova L, Le Goff X, Rasmussen HH, Cheperegin S, Drugeon G, Kress M, Arman I, Haenni AL, Celis JE, Philippe M, et al; Medline: 95082951 A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor" [see comments] Nature 1994;372:701-703.
[2] Drugeon G, Jean-Jean O, Frolova L, Le Goff X, Philippe M, Kisselev L, Haenni AL; Medline: 97315314 Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA." Nucleic Acids Res 1997;25:2254-2258.

871. (Ribosomal_L14e)Ribosomal protein L14
This family includes the eukaryotic ribosomal protein L14.
Number of members: 15

872. (Ribosomal_S27)
Ribosomal protein S27a

This family of ribosomal proteins consists mainly of the 40S ribosomal protein S27a which is synthesized as a C-terminal extension of ubiquitin (CEP). The S27a domain compromises the C-terminal half of the protein. The synthesis of ribosomal proteins as extensions of ubiquitin promotes their incorporation into nascent ribosomes by a transient metabolic stabilization and is required for efficient ribosome biogenesis [3]. The ribosomal extension protein S27a contains a basic region that is proposed to form a zinc finger; its fusion gene is proposed as a mechanism to maintain a fixed ratio between ubiquitin necessary for degrading proteins and ribosomes a source of proteins [2].

Number of members: 36

873. (Spermine_synth)
Spermine/spermidine synthase

Spermine and spermidine are polyamines. This family includes spermidine synthase that catalyses the fifth (last) step in the biosynthesis of spermidine from arginine, and spermine synthase.

Number of members: 39

[1] Mezquita J, Pau M, Mezquita C; Medline: 97449308 Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids." Gene 1997;195:313-319.
[2] Redman KL, Rechsteiner M; Medline: 89181932 Identification of the long ubiquitin extension as ribosomal protein S27a." Nature 1989;338:438-440.
[3] Finley D, Bartel B, Varshavsky A; Medline: 89181925 The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis." Nature 1989;338:394-401.

874. (Surp)

Surp module

[1] Denhez F, Lafyatis R; Medline: 94266805 Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot." J Biol Chem 1994;269:16170-16179.

This domain is also known as the SWAP domain. SWAP stands for Suppressor-of-White-APricot. It has been suggested that these domains may be RNA binding [1].

Number of members: 32

875. (TFIIE)

TFIIE alpha subunit

The general transcription factor TFIIE has an essential role in eukaryotic transcription initiation together with RNA polymerase II and other general factors. Human TFIIE consists of two subunits TFIIE-alpha Swiss:P29083 and TFIIE-beta Swiss:P29084 and joins the preinitiation complex after RNA polymerase II and TFIIF [1]. This family consists of the conserved amino terminal region of eukaryotic TFIIE-alpha [2] and proteins from archaebacteria that are presumed to be TFIIE-alpha subunits also Swiss:O29501 [3].

Number of members: 12

[1] Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder RG; Medline: 92065982 Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE." Nature 1991;354:398-401.

[2] Ohkuma Y, Hashimoto S, Roeder RG, Horikoshi M; Medline: 93087200 Identification of two large subdomains in TFIIE-alpha on the basis of homology between Xenopus and human sequences. Nucleic Acids Res 1992;20:5838-5838.

[3] Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA,

McKenney K, Adams MD, Loftus B, Venter JC, et al; Medline: 98049343 The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon Archaeoglobus fulgidus." Nature 1997;390:364-370.

876. (Transglut_core)

Cross-reference(s) PS00547; TRANSGLUTAMINASES

Transglutaminases (EC 2.3.2.13) (TGase) [1,2] are calcium-dependent enzymes that catalyze the cross-linking of proteins by promoting the formation of isopeptide bonds between the gamma-carboxyl group of a glutamine in one polypeptide chain and the epsilon-amino group of a lysine in a second polypeptide chain. TGases also catalyze the conjugation of polyamines to proteins. The best known transglutaminase is blood coagulation factor XIII, a plasma tetrameric protein composed of two catalytic A subunits and two non-catalytic B subunits. Factor XIII is responsible for cross-linking fibrin chains, thus stabilizing the fibrin clot. Other forms of transglutaminases are widely distributed in various organs, tissues and body fluids. Sequence data is available for the following forms of TGase:

- Transglutaminase K (Tgase K), a membrane-bound enzyme found in mammalian epidermis and important for the formation of the cornified cell envelope (gene TGM1).
- Tissue transglutaminase (TGase C), a monomeric ubiquitous enzyme located in the cytoplasm (gene TGM2).
- Transglutaminase 3, responsible for the later stages of cell envelope formation in the epidermis and the hair follicle (gene TGM3).
- Transglutaminase 4 (gene TGM4).

A conserved cysteine is known to be involved in the catalytic mechanism of TGases. The erythrocyte membrane band 4.2 protein, which probably plays an important role in regulating the shape of erythrocytes and their mechanical properties, is evolutionary related to TGases. However the active site cysteine is substituted by an alanine and the 4.2 protein does not show TGase activity.

Consensus pattern:[GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G [The first C is the active site residue] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ichinose A., Bottenus R.E., Davie E.W. J. Biol. Chem. 265:13411-13414(1990).
[2] Greenberg C.S., Birckbichler P.J., Rice R.H. FASEB J. 5:3071-3077(1991).

877. (TruB_N)
TruB family pseudouridylate synthase (N terminal domain)

Members of this family are involved in modifying bases in RNA molecules. They carry out the conversion of uracil bases to pseudouridine. This family includes TruB, a pseudouridylate synthase that specifically converts uracil 55 to pseudouridine in most tRNAs. This family also includes Cbf5p that modifies rRNA [2].

Number of members: 33

[1] Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J; Medline: 96079944 Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli." RNA 1995;1:102-112.
[2] Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D; Medline: 98139521 The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase." Genes Dev 1998;12:527-537.

878. (UDPGP)
UTP--glucose-1-phosphate uridylyltransferase

This family consists of UTP--glucose-1-phosphate uridylyltransferases, EC:2.7.7.9. Also known as UDP-glucose pyrophosphorylase (UDPGP) and Glucose-1-phosphate uridylyltransferase. UTP--glucose-1-phosphate uridylyltransferase catalyses the interconversion of MgUTP + glucose-1-phosphate and UDP-glucose + MgPPi [1]. UDP-glucose is an important intermediate in mammalian carbohydrate interconversion involved in

various metabolic roles depending on tissue type [1]. In Dictyostelium (slime mold) mutants in this enzyme abort the development cycle [2]. Also within the family is UDP-N-acetylglucosamine Swiss:Q16222 or AGX1 [3] and two hypothetical proteins from Borrelia burgdorferi the lyme disease spirochaete Swiss:O51893 and Swiss:O51036.

Number of members: 18

[1] Duggleby RG, Chao YC, Huang JG, Peng HL, Chang HY; Medline: 96202932 Sequence differences between human muscle and liver cDNAs for UDPglucose pyrophosphorylase and kinetic properties of the recombinant enzymes expressed in Escherichia coli." Eur J Biochem 1996;235:173-179.

[2] Ragheb JA, Dottin RP; Medline: 87231075 Structure and sequence of a UDP glucose pyrophosphorylase gene of Dictyostelium discoideum." Nucleic Acids Res 1987;15:3891-3906.

[3] Mio T, Yabe T, Arisawa M, Yamada-Okabe H; Medline: 98269105 The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression, and catalytic mechanism. J Biol Chem 1998;273:14392-14397.

879. (UPF004)
Uncharacterized protein family UPF0044 signature
Cross-reference(s) PS01301; UPF0044

The following uncharacterized proteins have been shown [1] to be highlysimilar:

- Bacillus subtilis hypothetical protein yqeI.
- Escherichia coli hypothetical protein yhbY and HI1333, the corresponding Haemophilus influenzae protein.
- Methanococcus jannaschii hypothetical protein MJ0652.

These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

Consensus pattern: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)- [LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

880. (zf-A20)

A20-like zinc finger

A20- (an inhibitor of cell death)-like zinc fingers. The zinc finger mediates self-association in A20. These fingers also mediate IL-1-induced NF-kappa B activation.

Number of members: 22

[1] Heyninck K, Beyaert R; Medline: 99126071 The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF- kappaB activation at the level of TRAF6. FEBS Lett 1999;442:147-150.

[2] De Valck D, Heyninck K, Van Criekinge W, Contreras R,Beyaert R, Fiers W; Medline: 96390831 A20, an inhibitor of cell death, self-associates by its zinc finger domain." FEBS Lett 1996;384:61-64.

[3] Song HY, Rothe M, Goeddel DV; Medline: 96270609 The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. Proc Natl Acad Sci U S A 1996;93:6721-6725.

[4] Opipari AW Jr, Boguski MS, Dixit VM; Medline: 90368626 The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein." J Biol Chem 1990;265:14705-14708.

881. (zf-PARP)

Poly(ADP-ribose) polymerase zinc finger domain

Cross-reference(s) PS00347; PARP_ZN_FINGER_1 PS50064; PARP_ZN_FINGER_2

Poly(ADP-ribose) polymerase (EC 2.4.2.30) (PARP) [1,2] is a eukaryotic enzyme that catalyzes the covalent attachment of ADP-ribose units from NAD(+) to various nuclear

acceptor proteins. This post-translational modification of nuclear proteins is dependent on DNA. It appears to be involved in the regulation of various important cellular processes such as differentiation, proliferation and tumor transformation as well as in the regulation of the molecular events involved in the recovery of the cell from DNA damage. Structurally, PARP, about 1000 amino-acids residues long, consists of three distinct domains: an N-terminal zinc-dependent DNA-binding domain, a central automodification domain and a C-terminal NAD-binding domain. The DNA-binding region contains a pair of zinc finger domains which have been shown to bind DNA in a zinc-dependent manner. The zinc finger domains of PARP seem to bind specifically to single-stranded DNA. DNA ligase III [3] contains, in its N-terminal section, a single copy of a zinc finger highly similar to those of PARP.

Consensus pattern: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C [The three C's and the H are zinc ligands] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROTNONE.

Note: This documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

[1] Althaus F.R., Richter C.R. Mol. Biol. Biochem. Biophys. 37:1-126(1987).
[2] de Murcia G., Menissier de Murcia J. Trends Biochem. Sci. 19:172-176(1994).
[3] Wei Y.-F., Robins P., Carter K., Caldecott K., Pappin D.J.C., Yu G.-L., Wang R.-P., Shell B.K., Nash R.A., Schar P., Barnes D.E., Haseltine W.A., Lindahl T. Mol. Cell. Biol. 15:3206-3216(1995).

882. Adenylylsulfate kinase (APS_kinase)

Enzyme that catalyses the phosphorylation of adenylylsulfate to 3'-phosphoadenylylsulfate. This domain contains an ATP binding P-loop motif. Number of members: 34

[1] MacRae IJ, Rose AB, Segel IH; Medline: 99003196 Adenosine 5'-phosphosulfate kinase from Penicillium chrysogenum. site- directed mutagenesis at putative phosphoryl-accepting and ATP P-loop residues. J Biol Chem 1998;273:28583-28589.

883. DNA polymerase family B signature DNA_POLYMERASE_B (DNA_pol_B)

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarity, a number of DNA polymerases have been grouped [1 to 7] under the designation of DNA polymerase family B. These are:

- Higher eukaryotes polymerases alpha.
- Higher eukaryotes polymerases delta.
- Yeast polymerase I/alpha (gene POL1), polymerase II/epsilon (gene POL2), polymerase III/delta (gene POL3) and polymerase REV3.
- Escherichia coli polymerase II (gene dinA or polB).
- Archaebacterial polymerases.
- Polymerases of viruses from the herpesviridae family.
- Polymerases from Adenoviruses.
- Polymerases from Baculoviruses.
- Polymerases from Chlorella viruses.
- Polymerases from Poxviruses.
- Bacteriophage T4 polymerase.
- Podoviridae bacteriophages Phi-29, M2 and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
- Polymerases encoded on mitochondrial linear DNA plasmids in various fungi and plants (Kluyveromyces lactis pGKL1 and pGKL2, Agaricus bitorquis pEM, Ascobolus immersus pAI2, Claviceps purpurea pCLK1, Neurospora Kalilo and Maranhar, maize S-1, etc).

Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. Its function is not yet known. However, it has been suggested [3] that it may be involved in binding a magnesium ion. This conserved region was selected as a signature for this family of DNA polymerases.

Consensus pattern [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC]
Sequences known to belong to this class detected by the patternALL, except for yeast polymerase II/epsilon, Agaricus bitorquis pEM and Sulfolobus solfataricus polymerase II.

[1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J. Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).
[2] Bernad A., Zaballos A., Salas M., Blanco L. EMBO J. 6:4219-4225(1987).
[3] Argos P. Nucleic Acids Res. 16:9909-9916(1988).
[4] Wang T.S.-F., Wong S.W., Korn D. FASEB J. 3:14-21(1989).
[5] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Eng. 3:461-467(1990).
[6] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
[7] Braithwaite D.K., Ito J. Nucleic Acids Res. 21:787-802(1993).

884. DNA polymerase family X signature - DNA_POLYMERASE_X (DNA_polymeraseX)

DNA polymerases (EC 2.7.7.7) can be classified, on the basis of sequence similarity [1], into at least four different groups: A, B, C and X. DNA polymerases that belong to family X are listed below [2]:

- Vertebrate polymerase beta, involved in DNA repair.
- Yeast polymerase IV (POL4) [3], an enzyme with similar characteristics to that of the mammalian polymerase beta.
- Terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31). TdT catalyzes the elongation of polydeoxynucleotide chains by terminal addition. One of the functions of this enzyme is the addition of nucleotides at the junction of rearranged Ig heavy chain and T cell receptor gene segments during the maturation of B and T cells.
- African Swine Fever virus protein O174L [4].
- Fission yeast hypothetical protein SpAC2F7.06c.

These enzymes are small (about 40 Kd) compared with other polymerases and their reaction mechanism operates via a distributive mode, i.e. they dissociate from the template-primer after addition of each nucleotide.

As a signature pattern for this family of DNA polymerases, a highly conserved region that contains a conserved arginine and two conserved aspartic acid residues were selected. The latter together with the arginine have been shown [5] to be involved in primer binding in polymerase beta.

Consensus pattern G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D- [LIVMFY](3)-x(2)-[SAP] Sequences known to belong to this class detected by the patternALL.

[1] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).

[2] Matsukage A., Nishikawa K., Ooi T., Seto Y., Yamaguchi M. J. Biol. Chem. 262:8960-8962(1987).

[3] Prasad R., Widen S.G., Singhal R.K., Watkins J., Prakash L., Wilson S.H. Nucleic Acids Res. 21:5301-5307(1993).

[4] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).

[5] Date T., Yamamoto S., Tanihara K., Nishimoto Y., Matsukage A. Biochemistry 30:5286-5292(1991).

885. DUF14 - Domain of unknown function

This domain is found in glutamate synthase, tungsten formylmethanofuran dehydrogenase subunit c (FwdC) and molybdenum formylmethanofuran dehydrogenase subunit c (FmdC). It has no known function. Number of members: 52

[1] Hochheimer A, Hedderich R, Thauer RK; Medline: 99035764. The formylmethanofuran dehydrogenase isoenzymes in Methanobacterium wolfei and Methanobacterium thermoautotrophicum: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme." Arch Microbiol 1998;170:389-393.

886. DUF18-Domain of unknown function

This domain of unknown function is found in several C. elegans proteins. The domain is 120 amino acids long and rich in cysteine residues. There are 16 conserved cysteine positions in the domain. Number of members: 34

887. DUF27-Domain of unknown function

This domain is found in a number of otherwise unrelated proteins. This domain is found at the C-terminus of the macro-H2A histone protein Swiss:Q02874. This domain is found in the non-structural proteins of several types of ssRNA viruses such as NSP2 from alphaviruses Swiss:P03317. This domain is also found on its own in a family of proteins from bacteria Swiss:P75918, archaebacteria Swiss:O59182 and eukaryotes Swiss:Q17432, suggesting that it is involved in an important and ubiquitous cellular process. Number of members: 66

888. DUF37-Domain of unknown function

This domain is found in short (70 amino acid) hypothetical proteins from various bacteria. The domain contains three conserved cysteine residues. Swiss:Q44066 from Aeromonas hydrophila has been found to have hemolytic activity (unpublished). Number of members: 19

889. EGF-like domain signatures. (EGF-like)

A sequence of about thirty to forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown [1 to 6] to be present, in a more or less conserved form, in a large number of other, mostly animal proteins. The proteins currently known to contain one or more copies of an EGF-like pattern are listed below.

- Adipocyte differentiation inhibitor (gene PREF-1) from mouse (6 copies).
- Agrin, a basal lamina protein that causes the aggregation of acetylcholine receptors on cultured muscle fibers (4 copies).
- Amphiregulin, a growth factor (1 copy).
- Betacellulin, a growth factor (1 copy).
- Blastula proteins BP10 and Span from sea urchin which are thought to be involved in pattern formation (1 copy).
- BM86, a glycoprotein antigen of cattle tick (7 copies).
- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity (1-2 copies). Homologous proteins are found in sea urchin - suBMP (1 copy) - and in Drosophila - the dorsal-ventral patterning protein tolloid (2 copies).
- Caenorhabditis elegans developmental proteins lin-12 (13 copies) and glp-1 (10 copies).
- Caenorhabditis elegans APX-1 protein, a patterning protein (4.5 copies).
- Calcium-dependent serine proteinase (CASP) which degrades the extracellular matrix proteins type I and IV collagen and fibronectin (1 copy).

- Cartilage matrix protein CMP (1 copy).
- Cartilage oligomeric matrix protein COMP (4 copies).
- Cell surface antigen 114/A10 (3 copies).
- Cell surface glycoprotein complex transmembrane subunit ASGP-2 from rat (2 copies).
- Coagulation associated proteins C, Z (2 copies) and S (4 copies).
- Coagulation factors VII, IX, X and XII (2 copies).
- Complement C1r components (1 copy).
- Complement C1s components (1 copy).
- Complement-activating component of Ra-reactive factor (RARF) (1 copy).
- Complement components C6, C7, C8 alpha and beta chains, and C9 (1 copy).
- Crumbs, an epithelial development protein from Drosophila (29 copies).
- Epidermal growth factor precursor (7-9 copies).
- Exogastrula-inducing peptides A, C, D and X from sea urchin (1 copy).
- Fat protein, a Drosophila cadherin-related tumor suppressor (5 copies).
- Fetal antigen 1, a probable neuroendocrine differentiation protein, which is derived from the delta-like protein (DLK) (6 copies).
- Fibrillin 1 (47 copies) and fibrillin 2 (14 copies).
- Fibropellins IA (21 copies), IB (13 copies), IC (8 copies), II (4 copies) and III (8 copies) from the apical lamina - a component of the extracellular matrix - of sea urchin.
- Fibulin-1 and -2, two extracellular matrix proteins (9-11 copies).
- Giant-lens protein (protein Argos), which regulates cell determination and axon guidance in the Drosophila eye (1 copy).
- Growth factor-related proteins from various poxviruses (1 copy).
- Gurken protein, a Drosophila developmental protein (1 copy).
- Heparin-binding EGF-like growth factor (HB-EGF), transforming growth factor alpha (TGF-alpha), growth factors Lin-3 and Spitz (1 copy); the precursors are membrane proteins, the mature form is located extracellular.
- Hepatocyte growth factor (HGF) activator (EC 3.4.21.-) (2 copies).
- LDL and VLDL receptors, which bind and transport low-density lipoproteins and very low-density lipoproteins (3 copies).
- LDL receptor-related protein (LRP), which may act as a receptor for endocytosis of extracellular ligands (22 copies).
- Leucocyte antigen CD97 (3 copies), cell surface glycoprotein EMR1 (6 copies) and cell surface glycoprotein F4/80 (7 copies).

- Limulus clotting factor C, which is involved in hemostasis and host defense mechanisms in japanese horseshoe crab (1 copy).

- Meprin A alpha subunit, a mammalian membrane-bound endopeptidase (1 copy).

- Milk fat globule-EGF factor 8 (MFG-E8) from mouse (2 copies).

- Neuregulin GGF-I and GGF-II, two human glial growth factors (1 copy).

- Neurexins from mammals (3 copies).

- Neurogenic proteins Notch, Xotch and the human homolog Tan-1 (36 copies), Delta (9 copies) and the similar differentiation proteins Lag-2 from Caenorhabditis elegans (2 copies), Serrate (14 copies) and Slit (7 copies) from Drosophila.

- Nidogen (also called entactin), a basement membrane protein from chordates (2-6 copies).

- Ookinete surface proteins (24 Kd, 25 Kd, 28 Kd) from Plasmodium (4 copies).

- Pancreatic secretory granule membrane major glycoprotein GP2 (1 copy).

- Perforin, which lyses non-specifically a variety of target cells (1 copy).

- Proteoglycans aggrecan (1 copy), versican (2 copies), perlecan (at least 2 copies), brevican (1 copy) and chondroitin sulfate proteoglycan (gene PG-M) (2 copies).

- Prostaglandin G/H synthase 1 and 2 (EC 1.14.99.1) (1 copy), which is found in the endoplasmatic reticulum.

- S1-5, a human extracellular protein whose ultimate activity is probably modulated by the environment (5 copies).

- Schwannoma-derived growth factor (SDGF), an autocrine growth factor as well as a mitogen for different target cells (1 copy).

- Selectins. Cell adhesion proteins such as ELAM-1 (E-selectin), GMP-140 (P-selectin), or the lymph-node homing receptor (L-selectin) (1 copy).

- Serine/threonine-protein kinase homolog (gene Pro25) from Arabidopsis thaliana, which may be involved in assembly or regulation of light-harvesting chlorophyll A/B protein (2 copies).

- Sperm-egg fusion proteins PH-30 alpha and beta from guinea pig (1 copy).

- Stromal cell derived protein-1 (SCP-1) from mouse (6 copies).

- TDGF-1, human teratocarcinoma-derived growth factor 1 (1 copy).

- Tenascin (or neuronectin), an extracellular matrix protein from mammals (14.5 copies), chicken (TEN-A) (13.5 copies) and the related proteins human tenascin-X (18 copies) and tenascin-like proteins TEN-A and TEN-M from Drosophila (8 copies).

- Thrombomodulin (fetomodulin), which together with thrombin activates protein C (6 copies).

- Thrombospondin 1, 2 (3 copies), 3 and 4 (4 copies), adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions.
- Thyroid peroxidase 1 and 2 (EC 1.11.1.8) from human (1 copy).
- Transforming growth factor beta-1 binding protein (TGF-B1-BP) (16 or 18 copies).
- Tyrosine-protein kinase receptors Tek and Tie (EC 2.7.1.112) (3 copies).
- Urokinase-type plasminogen activator (EC 3.4.21.73) (UPA) and tissue plasminogen activator (EC 3.4.21.68) (TPA) (1 copy).
- Uromodulin (Tamm-horsfall urinary glycoprotein) (THP) (3 copies).
- Vitamin K-dependent anticoagulants protein C (2 copies) and protein S (4 copies) and the similar protein Z, a single-chain plasma glycoprotein of unknown function (2 copies).
- 63 Kd sperm flagellar membrane protein from sea urchin (3 copies).
- 93 Kd protein (gene nel) from chicken (5 copies).
- Hypothetical 337.6 Kd protein T20G5.3 from Caenorhabditis elegans (44 copies).

The functional significance of EGF domains in what appear to be unrelated proteins is not yet clear. However, a common feature is that these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines strongly vary in length as shown in the following schematic representation of the EGF-like domain:

```
          +------------------+      +------------------------+          |            |    |
| x(4)-C-x(0,48)-C-x(3,12)-C-x(1,70)-C-x(1,6)-C-x(2)-G-a-x(0,21)-G-x(2)-C-x     |
|       ***********************************
     +------------------+
```

'C': conserved cysteine involved in a disulfide bond.

'G': often conserved glycine

'a': often conserved aromatic amino acid

'*': position of both patterns.

'x': any residue

The region between the 5th and 6th cysteine contains two conserved glycines of which at least one is present in most EGF-like domains. Two patterns were created for this domain, each including one of these C-terminal conserved glycine residues.

Consensus pattern: C-x-C-x(5)-G-x(2)-C [The 3 C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern A majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT87 proteins, of which 27 can be considered as possible candidates.

Consensus pattern: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C [The three C's are involved in disulfide bonds]Sequences known to belong to this class detected by the patternA majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT83 proteins, of which 49 can be considered as possible candidates. Note The beta chain of the integrin family of proteins contains 2 cysteine- rich repeats which were said to be dissimilar with the EGF pattern [7].

Note Laminin EGF-like repeats (see <PDOC00961>) are longer than the average EGF module and contain a further disulfide bond C-terminal of the EGF-like region. Perlecan and agrin contain both EGF-like domains and laminin-type EGF-like domains. Note the pattern do not detect all of the repeats of proteins with multiple EGF-like repeats. Note see <PDOC00913> for an entry describing specifically the subset of EGF- like domains that bind calcium.

[1] Davis C.G. New Biol. 2:410-419(1990).

[2] Blomquist M.C., Hunt L.T., Barker W.C. Proc. Natl. Acad. Sci. U.S.A. 81:7363-7367(1984).

[3] Barker W.C., Johnson G.C., Hunt L.T., George D.G. Protein Nucl. Acid Enz. 29:54-68(1986).

[4] Doolittle R.F., Feng D.F., Johnson M.S. Nature 307:558-560(1984).

[5] Appella E., Weber I.T., Blasi F. FEBS Lett. 231:1-4(1988).

[6] Campbell I.D., Bork P. Curr. Opin. Struct. Biol. 3:385-392(1993).

[7] Tamkun J.W., DeSimone D.W., Fonda D., Patel R.S., Buck C., Horwitz A.F., Hynes R.O. Cell 46:271-282(1986).

890. Ham1 family (Ham1p_like)

This family consists of the HAM1 protein Swiss:P47119 and hypothetical archaeal bacterial and C. elegans proteins. HAM1 controls 6-N-hydroxylaminopurine (HAP) sensitivity and mutagenesis in S. cerevisiae Swiss:P47119 [1]. The HAM1 protein protects the cell from HAP, either on the level of deoxynucleoside triphosphate or the DNA level by a yet unidentified set of reactions [1]. Number of members: 19

[1] Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H, Pavlov YI; Medline: 96381244 HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast Saccharomyces cerevisiae." Yeast 1996;12:17-29.

891. (HCO3_cotransp)

Anion exchange is a cellular transport function which contributes to the regulation of cell pH and volume. Anion exchangers are a family of functionally related proteins that contributes to these properties by maintaining the intracellular level of the two principal anions: chloride and HCO3-. The best characterized anion exchanger is the band 3 protein [1], which is an erythrocyte anion exchange membrane glycoprotein. Band 3 is a protein of about 900 amino acids which consists of a cytoplasmic N-terminal domain of about 400 residues and an hydrophobic C-terminal section of about 500 residues that contains at least ten transmembrane regions. The cytoplasmic domain provides binding sites for cytoskeletal proteins, while the integral membrane domain is responsible for anion transport. Band 3 protein is specific to erythroid cells, at least two other proteins [2] structurally and functionally related to band 3, are found in nonerythroid tissues:

- AE2 (or B3 related protein; B3RP), a protein of 1200 residues, which seems to be present in a variety of cell types including lymphoid, kidney, and choroid plexus.
- AE3, a protein of 1200 residues, which is specific to neurons.

Structurally AE2 and AE3 are very similar to band 3, the main difference being an extension of some 300 residues of the N-terminal domain in AE2 and AE3.

Two signature patterns were developed for these proteins. The first pattern is based on a conserved stretch of sequence that contains four clustered positive charged residues and which is located at the C-terminal extremity of the cytoplasmic domain, just before the first transmembrane segment from the integral domain. The second pattern is based on the perfectly conserved sequence of the fifth transmembrane segment; this segment contains a lysine, which is the covalent binding site for the isothiocyanate group of DIDS, an inhibitor of anion exchange.

Consensus pattern F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L Sequences known to belong to this class detected by the pattern ALL.

[1] Jay D., Cantley L. Annu. Rev. Biochem. 55:511-538(1986).
[2] Reithmeier R.A.F. Curr. Opin. Struct. Biol. 3:515-523(1993).

892. ATP phosphoribosyltransferase signature (HisG)

ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern a region located in the C-terminal part of this enzyme was selected.

Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM] Sequences known to belong to this class detected by the pattern ALL.

893. HNH endonuclease (HNH)

Number of members: 56

[1] Shub DA, Goodrich-Blair H, Eddy SR; Medline: 95117127 Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns." Trends Biochem Sci 1994;19:402-404.
[2] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site- specific endonucleases

and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

[3] Gorbalenya AE; Medline: 95004046 Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family." Protein Sci 1994;3:1117-1120.

894. NEUROHYPOPHYS_HORM (hormone5)

Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels. Like the majority of active peptides, both hormones are synthesized as larger protein precursors that are enzymatically converted to their mature forms. Peptides belonging to this family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin, glumitocin, aspargtocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin), octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs (conopressins G and S) [2]. The pattern developed to detect this category of peptides spans their entire sequence and includes four invariant amino acid residues.

Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide bond]. Sequences known to belong to this class detected by the pattern ALL.

[1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988).

[2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein Res. 45:482-487(1995).

895. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK)

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents such as trimethoprim or sulfonamides. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine

pyrophosphate. This is the first step in a three-step pathway leading to 7,8-dihydrofolate. Bacterial HPPK (gene folK or sulD) [1] is a protein of 160 to 270 amino acids. In the lower eukaryote Pneumocystis carinii, HPPK is the central domain of a multifunctional folate synthesis enzyme (gene fas) [2]. As a signature for HPPK, a conserved region located in the central section of these enzymes was selected.

Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992).

[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).

896. Metalloenzyme superfamily (Metalloenzyme)

This family includes phosphopentomutase Swiss:P07651 and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, Swiss:P37689. This family is also related to alk_phosphatase [1]. The alignment contains the most conserved residues that are probably involved in metal binding and catalysis. Number of members: 34

[1] Galperin MY, Bairoch A, Koonin EV; Medline: 99180418 A superfamily of metalloenzymes unifies phosphopentomutase and cofactor- independent phosphoglycerate mutase with alkaline phosphatases and sulfatases." Protein Sci 1998;7:1829-1835.

897. Penicillin amidase (Penicil_amidase)

Penicillin amidase or penicillin acylase EC:3.5.1.11 catalyses the hydrolysis of benzylpenicillin to phenylacetic acid and 6-aminopenicillanic acid (6-APA) a key intermediate in the the synthesis of penicillins [1]. Also in the family is cephalosporin acylase Swiss:P07662 and Swiss:P29958 aculeacin A acylase which are involved in the synthesis of related peptide antibiotics. Number of members: 13

[1] Verhaert RM, Riemens AM, van der Laan JM, van Duin J, Quax WJ; Medline: 97438505 Molecular cloning and analysis of the gene encoding the thermostable penicillin G acylase from Alcaligenes faecalis. Appl Environ Microbiol 1997;63:3412-3418.
[2] Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; Medline: 95115804 Penicillin acylase has a single-amino-acid catalytic centre." Nature 1995;373:264-268.

898. Phosphoribosyl-AMP cyclohydrolase (PRA-CH)
This enzyme catalyses the third step in the histidine biosynthetic pathway. It requires Zn ions for activity. Number of members: 13

[1] D'Ordine RL, Klem TJ, Davisson VJ; Medline: 99129952 N1-(5'-phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: purification and characterization of a unique metalloenzyme. Biochemistry 1999;38:1537-1546.

899. Phosphoribosyl-ATP pyrophosphohydrolase (PRA-PH)
This enzyme catalyses the second step in the histidine biosynthetic pathway. Number of members: 32

[1] Keesey JK Jr, Bigelis R, Fink GR; Medline: 79216449 The product of the his4 gene cluster in Saccharomyces cerevisiae. A trifunctional polypeptide." J Biol Chem 1979 Aug 10;254:7427-7433.
[2] Bruni CB, Carlomagno MS, Formisano S, Paolella G; Medline: 86310274 Primary and secondary structural homologies between the HIS4 gene product of Saccharomyces cerevisiae and the hisIE and hisD gene products of Escherichia coli and Salmonella typhimurium." Mol Gen Genet 1986;203:389-396.

900. Prokaryotic membrane lipoprotein lipid attachment site (PstS)
In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a

glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- Escherichia coli lipoprotein-34 (gene nlpB).
- Escherichia coli lipoprotein nlpC.
- Escherichia coli lipoprotein nlpD.
- Escherichia coli osmotically inducible lipoprotein B (gene osmB).
- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
- Escherichia coli copper homeostasis protein cutF (or nlpE).
- Escherichia coli plasmids traT proteins.
- Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
- Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
- Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- Fibrobacter succinogenes endoglucanase cel-3.
- Haemophilus influenzae proteins Pal and Pcp.
- Klebsiella pullulunase (gene pulA).
- Klebsiella pullulunase secretion protein pulS.
- Mycoplasma hyorhinis protein p37.
- Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC).
- Neisseria outer membrane protein H.8.
- Pseudomonas aeruginosa lipopeptide (gene lppL).
- Pseudomonas solanacearum endoglucanase egl.
- Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- Rickettsia 17 Kd antigen.
- Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
- Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
- Treponema pallidium 34 Kd antigen.
- Treponema pallidium membrane protein A (gene tmpA).

- Vibrio harveyi chitobiase (gene chb).
- Yersinia virulence plasmid protein yscJ.
- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper-binding protein. This is the first archaebacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, a consensus pattern was derived and a set of rules to identify this type of post-translational modification.

Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTsome 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).
[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).
[3] von Heijne G. Protein Eng. 2:531-534(1989).
[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

901. Ribosome recycling factor (RRF)

The ribosome recycling factor (RRF / ribosome release factor) dissociates the ribosome from the mRNA after termination of translation, and is essential bacterial growth [1]. Thus ribosomes are "recycled" and ready for another round of protein synthesis. Number of members: 27

[1] Janosi L, Shimizu I, Kaji A; Medline: 94240115 Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth." Proc Natl Acad Sci U S A 1994;91:4249-4253.

902. S-layer homology(SLH)

S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain has been found in:

- S-layer glycoprotein of Acetogenium kivui (3 copies).
- S-layer 125 Kd protein of Bacillus sphaericus (3 copies).
- S-layer protein of Bacillus anthracis (3 copies).
- S-layer protein of Bacillus licheniformis (3 copies).
- S-layer protein (HWP) from Bacillus brevis strain HPD31 (3 copies).
- Middle cell wall protein (MWP) from Bacillus brevis strain 47 (3 copies).
- S-layer protein (p100) of Thermus thermophilus (1 copy).
- Outer membrane protein Omp-alpha from Thermotoga maritima (1 copy).
- Cellulosome anchoring protein (gene ancA), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from Clostridium thermocellum (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation).
- Amylopullulanase (gene amyB) from Thermoanaerobacter thermosulfurogenes (3 copies)
- Amylopullulanase (gene aapT) from Bacillus strain XAL-601 (3 copies).
- Endoglucanase from Bacillus strain KSM-635 (3 copies).
- Exoglucanase (gene xynX) from Clostridium thermocellum (3 copies).
- Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account).
- Protein involved in butirosin production (ButB) from Bacillus circulans (2 incomplete copies; 3 copies if three frameshifts are taken into account).
- Two hypothetical proteins from Synechocystis strain PCC 6803 (1 copy each).
- A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 copy; 3 copies if two frameshifts are taken into account).

SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible

to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length.

Consensus pattern[LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[FYWPDA]-x(4)-[LIV]-x(2)-[GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x- [PGAV]-x(3,10)-[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Beveridge T.J. Curr. Opin. Struct. Biol. 4:204-212(1994).
[2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S., Baumeister W. J. Bacteriol. 176:1224-1233(1994).
[3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).

903. Queuine tRNA-ribosyltransferase (TGT)

This is a family of queuine tRNA-ribosyltransferases EC:2.4.2.29, also known as tRNA-guanine transglycosylase and guanine insertion enzyme. Queuine tRNA-ribosyltransferase modifies tRNAs for asparagine, aspartic acid, histidine and tyrosine with queuine. It catalyses the exchange of guanine-34 at the wobble position with 7-aminomethyl-7-deazaguanine, and the addition of a cyclopentenediol moiety to 7-aminomethyl-7-deazaguanine-34 tRNA; giving a hypermodified base queuine in the wobble position [1,2].The aligned region contains a zinc binding motif C-x-C-x2-C-x29-H, and important tRNA and 7-aminomethyl-7deazaguanine binding residues [1]. Number of members: 27

[1] Romier C, Reuter K, Suck D, Ficner R; Medline: 96256303 Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange." EMBO J 1996;15:2850-2857.
[2] Garcia GA, Koch KA, Chong S; Medline: 93287116 tRNA-guanine transglycosylase from Escherichia coli. Overexpression, purification and quaternary structure." J Mol Biol 1993;231:489-497.

904. ThiC Family (ThiC)

ThiC is found within the thiamine biosynthesis operon. ThiC is involved in pyrimidine biosynthesis [2]. ThiC catalyzes the substitution of the pyrophosphate of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by 4-methyl-5-(beta-hydroxyethyl)thiazole phosphate to yield thiamine phosphate [3]. Number of members: 12

[1] Vander Horn PB, Backstrom AD, Stewart V, Begley TP; Medline: 93163063 Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in Escherichia coli K-12." J Bacteriol 1993;175:982-992.

[2] Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP, Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ, Xi J; Medline: 99311269 Thiamin biosynthesis in prokaryotes." Arch Microbiol 1999;171:293-300.

[3] Zhang Y, Taylor SV, Chiu HJ, Begley TP; Medline: 97284509 Characterization of the Bacillus subtilis thiC operon involved in thiamine biosynthesis." J Bacteriol 1997;179:3030-3035.

905. Putative tRNA binding domain (tRNA_bind)

This domain is found in prokaryotic methionyl-tRNA synthetases, prokaryotic phenylalanyl tRNA synthetases the yeast GU4 nucleic-binding protein (G4p1 or p42, ARC1) [2], human tyrosyl-tRNA synthetase [1], and endothelial-monocyte activating polypeptide II. G4p1 binds specifically to tRNA form a complex with methionyl-tRNA synthetases [2]. In human tyrosyl-tRNA synthetase this domain may direct tRNA to the active site of the enzyme [2]. This domain may perform a common function in tRNA aminoacylation [1]. Number of members: 12

[1] Kleeman TA, Wei D, Simpson KL, First EA; Medline: 97306356 Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine." J Biol Chem 1997;272:14420-14425.

[2] Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A, Mann M, Hurt EC; Medline: 97050848 The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl-and glutamyl-tRNA synthetases." EMBO J 1996;15:5437-5448.

906. UbiA prenyltransferase family signature (UbiA)

The following prenyltransferases are evolutionary related [1,2]:

- Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA).
- Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene COQ2).
- Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene COX10) and from bacteria (genes cyoE or ctaB).

These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and was used as a signature pattern.

Consensus pattern N-x(3)-[DE]-x(2)-[LIF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Melzer M., Heide L. Biochim. Biophys. Acta 1212:93-102(1994).
[2] Mogi T., Saiki K., Anraku Y. Mol. Microbiol. 14:391-398(1994).

907. Uncharacterized protein family UPF0044 signature (UPF0044)

The following uncharacterized proteins have been shown [1] to be highly similar:

- Bacillus subtilis hypothetical protein yqeI.
- Escherichia coli hypothetical protein yhbY and HI1333, the corresponding Haemophilus influenzae protein.
- Methanococcus jannaschii hypothetical protein MJ0652.

These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

Consensus pattern L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)- [LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the patternALL.

908. ATP synthase (C/AC39) subunit (vATP-synt_AC39)

This family includes the AC39 subunit from vacuolar ATP synthase Swiss:P32366 [1], and the C subunit from archaebacterial ATP synthase [2]. The family also includes subunit C

from the Sodium transporting ATP synthase from Enterococcus hirae Swiss:P43456 [3]. Number of members: 12

[1] Bauerle C, Ho MN, Lindorfer MA, Stevens TH; Medline: 93286119 The Saccharomyces cerevisiae VMA6 gene encodes the 36-kDa subunit of the vacuolar H(+)-ATPase membrane sector." J Biol Chem 1993;268:12749-12757.

[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

[3] Takase K, Kakinuma S, Yamato I, Konishi K, Igarashi K, Kakinuma Y; Medline: 94209269 Sequencing and characterization of the ntp gene cluster for vacuolar- type Na(+)-translocating ATPase of Enterococcus hirae." J Biol Chem 1994;269:11037-11044.

909. ATP synthase (E/31 kDa) subunit (vATP-synt_E)

This family includes the vacuolar ATP synthase E subunit [1], as well as the archaebacterial ATP synthase E subunit [2]. Number of members: 24

[1] Foury F; Medline: 91009356 The 31-kDa polypeptide is an essential subunit of the vacuolar ATPase in Saccharomyces cerevisiae." J Biol Chem 1990;265:18554-18560.

[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

910. (WW)

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro.

It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

- Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C- terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.
- Utrophin, a dystrophin-like protein of unknown function.
- Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].
- Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].
- Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>, followed by a histidine-rich region, 3 WW domains and a HECT domain.
- Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.
- Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).
- Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.
- IQGAP, a human GTPase activating protein acting on ras. It contains an N- terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

- Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2- type myosin, each containing two WW-domains at the N-terminus.
- Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.
- Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT8. Sequences known to belong to this class detected by the profileALL.

[1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).
[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).
[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).
[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).
[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).
[6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).

911. Xeroderma pigmentosum (XP) [1] (XPG_1)

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].

XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:

- Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9].
- Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease.

In addition to the proteins listed in the above groups, this family also includes:
- Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs.
- Yeast EXO1 (DHS1), a protein with probably the same function as exo1.
- Yeast DIN7.

Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset.

Two signature patterns were developed for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.

Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

[2] Scherly D., Nouspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993).

[3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993).

[4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994).

[5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994).

[6] Szankasi P., Smith G.R. Science 267:1166-1169(1995).

[7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993).

[8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).

[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).

912. 5-formyltetrahydrofolate cyclo-ligase (5-FTHF_cyc-lig)

5-formyltetrahydrofolate cyclo-ligase or methenyl-THF synthetase EC:6.3.3.2 catalyses the interchange of 5-formyltetrahydrofolate (5-FTHF) to 5-10-methenyltetrahydrofolate, this requires ATP and Mg2+ [1]. 5-FTHF is used in chemotherapy where it is clinically known as Leucovorin [2].

Number of members: 23

[1] Dayan A, Bertrand R, Beauchemin M, Chahla D, Mamo A, Filion M, Skup D, Massie B, Jolivet J; Medline: 96096540 Cloning and characterization of the human 5,10-methenyltetrahydrofolate synthetase-encoding cDNA." Gene 1995;165:307-311.

[2] Maras B, Stover P, Valiante S, Barra D, Schirch V; Medline: 94308074 Primary structure and tetrahydropteroylglutamate binding site of rabbit liver cytosolic 5,10-methenyltetrahydrofolate synthetase." J Biol Chem 1994;269:18429-18433.

913. Cytosolic long-chain acyl-CoA thioester hydrolase (Acyl-CoA_hydro)

This family consist of various cytosolic long-chain acyl-CoA thioester hydrolases including human and rat [1,2]. The aligned region is repeated with in the sequence of human and rat

cytosolic long-chain acyl-CoA thioester hydrolases of this family. Long-chain acyl-CoA hydrolases hydrolyse palmitoyl-CoA to CoA and palmitate, they also catalyse the hydrolysis of other long chain fatty acyl-CoA thioesters. Long-chain acyl-CoA hydrolases are present in all living organisms and they may provide a mechanism for the control of lipid metabolism [1].

Number of members: 24

[1]Yamada J, Furihata T, Iida N, Watanabe T, Hosokawa M, Satoh T, Someya A, Nagaoka I, Suga T; Medline: 97236308 Molecular cloning and expression of cDNAs encoding rat brain and liver cytosolic long-chain acyl-CoA hydrolases." Biochem Biophys Res Commun 1997;232:198-203.

[2] Broustas CG, Larkins LK, Uhler MD, Hajra AK; Medline: 96209964 Molecular cloning and expression of cDNA encoding rat brain cytosolic acyl-coenzyme A thioester hydrolase." J Biol Chem 1996;271:10470-10476.

914. Agglutinin

Lectin (probable mannose binding)

Members of this family are plant lectins. Many if not all are mannose specific.

Number of members: 87

[1] Wright CS, Hester G; Medline: 97094989 The 2.0 A structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes." Structure 1996;4:1339-1352.

915. (ANF_RECEPTORS)

Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger.

Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic

peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.

GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C- terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short.

A pattern was developed from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.

Consensus patternG-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Garbers D.L. New Biol. 2:499-504(1990).
[2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).

916. (Apocytochrome)
Cytochrome c family heme-binding site signature

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

Consensus patternC-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the patternALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT454.

Note: some cytochrome c's have more than a single bound heme groupc4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 !

[1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

917. ATP-synt_A-c. ATP synthase Alpha chain, C terminal

[1] Medline: 94344236. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Abrahams JP, Leslie AG, Lutter R, Walker JE; Nature 1994;370:621-628.

Number of members: 125

918. (Basic)

Myc-type, 'helix-loop-helix' dimerization domain signature

HELIX_LOOP_HELIX

A number of eukaryotic proteins, which probably are sequence specific DNA- binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are refered as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also refered to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW.

- The myc family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation.

- Proteins involved in myogenesis (the induction of muscle cells). In mammals MyoD1 (Myf-3), myogenin (Myf-4), Myf-5, and Myf-6 (Mrf4 or herculin), in birds CMD1 (QMF-1), in Xenopus MyoD and MF25, in Caenorhabditis elegans CeMyoD, and in Drosophila nautilus (nau).

- Vertebrate proteins that bind specific DNA sequences ('E boxes') in various immunoglobulin chains enhancers: E2A or ITF-1 (E12/pan-2 and E47/pan-1), ITF-2 (tcf4), TFE3, and TFEB.

- Vertebrate neurogenic differentiation factor 1 that acts as differentiation factor during neurogenesis.

- Vertebrate MAX protein, a transcription regulator that forms a sequence- specific DNA-binding protein complex with myc or mad.

- Vertebrate Max Interacting Protein 1 (MXI1 protein) which acts as a transcriptional repressor and may antagonize myc transcriptional activity by competing for max.

- Proteins of the bHLH/PAS superfamily which are transcriptional activators. In mammals, AH receptor nuclear translocator (ARNT), single-minded homologs (SIM1 and SIM2), hypoxia-inducible factor 1 alpha (HIF1A), AH receptor (AHR), neuronal pas domain proteins (NPAS1 and NPAS2), endothelial pas domain protein 1 (EPAS1), mouse ARNT2, and human BMAL1. In drosophila, single-minded (SIM), AH receptor nuclear translocator (ARNT), trachealess protein (TRH), and similar protein (SIMA).

- Mammalian transcription factors HES, which repress transcription by acting on two types of DNA sequences, the E box and the N box.

- Mammalian MAD protein (max dimerizer) which acts as transcriptional repressor and may antagonize myc transcriptional activity by competing for max.

- Mammalian Upstream Stimulatory Factor 1 and 2 (USF1 and USF2), which bind to a symmetrical DNA sequence that is found in a variety of viral and cellular promoters.

- Human lyl-1 protein; which is involved, by chromosomal translocation, in T- cell leukemia.

- Human transcription factor AP-4.

- Mouse helix-loop-helix proteins MATH-1 and MATH-2 which activate E box- dependent transcription in collaboration with E47.

- Mammalian stem cell protein (SCL) (also known as tal1), a protein which may play an important role in hemopoietic differentiation. SCL is involved, by chromosomal translocation, in stem-cell leukemia.

- Mammalian proteins Id1 to Id4 [5]. Id (inhibitor of DNA binding) proteins lack a basic DNA-binding domain but are able to form heterodimers with other HLH proteins, thereby inhibiting binding to DNA.

- Drosophila extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins.

- Human Sterol Regulatory Element Binding Protein 1 (SREBP-1), a transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) found in the flanking region of the LDLR gene and in other genes.

- Drosophila achaete-scute (AS-C) complex proteins T3 (l'sc), T4 (scute), T5 (achaete) and T8 (asense). The AS-C proteins are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system.

- Mammalian homologs of achaete-scute proteins, the MASH-1 and MASH-2 proteins.

- Drosophila atonal protein (ato) which is involved in neurogenesis.

- Drosophila daughterless (da) protein, which is essential for neurogenesis and sex-determination.

- Drosophila deadpan (dpn), a hairy-like protein involved in the functional differentiation of neurons.

- Drosophila delilah (dei) protein, which is plays an important role in the differentiation of epidermal cells into muscle.

- Drosophila hairy (h) protein, a transcriptional repressor which regulates the embryonic segmentation and adult bristle patterning.

- Drosophila enhancer of split proteins E(spl), that are hairy-like proteins active during neurogenesis. also act as transcriptional repressors.

- Drosophila twist (twi) protein, which is involved in the establishment of germ layers in embryos.

- Maize anthocyanin regulatory proteins R-S and LC.

- Yeast centromere-binding protein 1 (CPF1 or CBF1). This protein is involved in chromosomal segregation. It binds to a highly conserved DNA sequence, found in centromers and in several promoters.

- Yeast INO2 and INO4 proteins.

- Yeast phosphate system positive regulatory protein PHO4 which interacts with the upstream activating sequence of several acid phosphatase genes.

- Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression.

- Neurospora crassa nuc-1, a protein that activates the transcription of structural genes for phosphorus acquisition.
- Fission yeast protein esc1 which is involved in the sexual differentiation process.

The schematic representation of the helix-loop-helix domain is shown here:
xxxxxxxxxxxxxxxxxxxxxxxx------------------xxxxxxxxxxxxxxxxxxxxxxx Amphipathic helix 1 Loop Amphipathic helix 2

The signature pattern that had been developed to detect this domain spans completely the second amphipathic helix.

Consensus pattern[DENSTAP]-[KR]-[LIVMAGSNT]-{FYWCPHKR}-[LIVMT]-[LIVM]-x(2)-[STAV]-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]-{P}-{P}- [LIVMRKHQ]
Sequences known to belong to this class detected by the pattern the majority but far from all.
Other sequence(s) detected in SWISS-PROT135.

[1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).
[2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).
[3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).
[4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990).
[5] Riechmann V., van Cruechten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

919. (Beta-lactamase)
Beta-lactamases classes -A, -C, and -D active site

Beta-lactamases (EC 3.5.2.6) [1,2] are enzymes which catalyze the hydrolysis of an amide bond in the beta-lactam ring of antibiotics belonging to the penicillin/cephalosporin family. Four kinds of beta-lactamase have been identified [3]. Class-B enzymes are zinc containing proteins whilst class -A, C and D enzymes are serine hydrolases. The three classes of serine beta-
lactamases are evolutionary related and belong to a superfamily [4] that also includes DD-peptidases and a variety of other penicillin-binding proteins (PBP's). All these proteins contain a Ser-x-x-Lys motif, where the serine is the active site residue. Although clearly homologous, the sequences of the three classes of serine beta-lactamases exhibit a large

degree of variability and only a small number of residues are conserved in addition to the catalytic serine.

Since a pattern detecting all serine beta-lactamases would also pick up many unrelated sequences, it was decided to provide specific patterns, centered on the active site serine, for each of the three classes.

Consensus pattern [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC] [S is the active site residue] Sequences known to belong to this class detected by the patternALL class-A beta-lactamases. Other sequence(s) detected in SWISS-PROT7.

Consensus pattern F-E-[LIVM]-G-S-[LIVMG]-[SA]-K [The first S is the active site residue] Sequences known to belong to this class detected by the patternALL class-C beta-lactamases. Other sequence(s) detected in SWISS-PROTNONE.

Consensus pattern [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI] [S is the active site residue] Sequences known to belong to this class detected by the patternALL class-D beta-lactamases. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ambler R.P. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 289:321-331(1980).
[2] Pastor N., Pinero D., Valdes A.M., Soberon X. Mol. Microbiol. 4:1957-1965(1990).
[3] Bush K. Antimicrob. Agents Chemother. 33:259-263(1989).
[4] Joris B., Ghuysen J.-M., Dive G., Renard A., Dideberg O., Charlier P., Frere J.M., Kelly J.A., Boyington J.C., Moews P.C., Knox J.R. Biochem. J. 250:313-324(1988).

920. Biotin protein ligase (BPL)

Biotin is covalently attached at the active site of certain enzymes that transfer carbon dioxide from bicarbonate to organic acids to form cellular metabolites. Biotin protein ligase (BPL) is the enzyme responsible for attaching biotin to a specific lysine at the active site of biotin enzymes. Each organism probably has only one BPL. Biotin attachment is a two step reaction that results in the formation of an amide linkage between the carboxyl group of biotin and the epsilon-amino group of the modified lysine [2].
Number of members: 26

[1] Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW; Medline: 93028443 Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains." Proc Natl Acad Sci USA 1992;89:9257-9261.
[2] Chapman-Smith A, Cronan JE Jr; Medline: 10470036 The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity." Trends Biochem Sci 1999;24:359-363.

921. (BRCA2_repeat)

The alignment covers only the most conserved region of the repeat. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature

[1] Bork P, Blomberg N, Nilges M; Medline: 96241568 Internal repeats in the BRCA2 protein sequence." Nat Genet 1996;13:22-23.

Number of members: 63

922. (C6)

This domain of unknown function is found in the C. elegans protein Swiss:Q19522. It is presumed to be an extracellular domain. The C6 domain contains six conserved cysteine residues in most copies of the domain. However some copies of the domain are missing cysteine residues 1 and 3 suggesting that these form a disulphide bridge.
Number of members: 23

923. Cadherin cytoplasmic region (Cadherin_C_term)

Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn [1].
Number of members: 59

[1] Yap AS, Niessen CM, Gumbiner BM; Medline: 98234411 The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with p120ctn." J Cell Biol 1998;141:779-789.
[2] Barth AI, Nathke IS, Nelson WJ; Medline: 97471931 Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways." Curr Opin Cell Biol 1997;9:683-690.
[3] Braga VM, Machesky LM, Hall A, Hotchin NA; Medline: 97327766 The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts." J Cell Biol 1997;137:1421-1431.

924. Clathrin propeller repeat (Clathrin_propel)

Clathrin is the scaffold protein of the basket-like coat that surrounds coated vesicles. The soluble assembly unit, a triskelion, contains three heavy chains and three light chains in an extended three-legged structure. Each leg contains one heavy and one light chain. The N-terminus of the heavy chain is known as the globular domain, and is composed of seven repeats which form a beta propeller [1].
Number of members: 61

[1] ter Haar E, Musacchio A, Harrison SC, Kirchhausen T; Medline: 99043510 Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker." Cell. 1998;95:563-573.

925. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature (complex1_30Kd)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 30 Kd (in mammals) which has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in Neurospora crassa.

- Mitochondrial encoded in Paramecium (protein P1), and in the slime mold Dictyostelium discoideum (ORF 209).
- Chloroplast encoded in various higher plants (ORF 159). It is also present in bacteria:
- In the cyanobacteria Synechocystis strain PCC 6803 (gene ndhJ).
- Subunit C of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoC).
- Subunit NQO5 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.

This protein, in its mature form, consists of from 157 to 266 amino acid residues. The best conserved region is located in the C-terminal section and can be used as a signature pattern.

Consensus pattern E-R-E-x(2)-[DE]-[LIVMFY](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMYS] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
[2]Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

926. Respiratory-chain NADH dehydrogenase 49 Kd subunit signature (complex1_49Kd)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 49 Kd (in mammals), which is the third largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 49 Kd subunit has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in Neurospora crassa.
- Mitochondrial encoded in protozoan such as Paramecium (ORF 400), Leishmania and Trypanosoma (MURF 3).
- Chloroplast encoded in various higher plants (ORF 392).

The 49 Kd subunit is highly similar to [3,4]:

- Subunit D of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoD).

- Subunit NQO4 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
- Subunit 5 of Escherichia coli formate hydrogenlyase (gene hycE).
- Subunit G of Escherichia coli hydrogenase-4 (gene hyfG).

A highly conserved region was seleceted as signature pattern, located in the N-terminal section of this subunit.

Consensus pattern [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMTN]-x-E-x-[KRQ] Sequences known to belong to this class detected by the patternALL.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
[3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).
[4] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

927. (COX2)

Cytochrome c oxidase (EC 1.9.3.1) [1,2] is an oligomeric enzymatic complex which is a component of the respiratory chain and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. The enzyme complex consists of 3-4 subunits (prokaryotes) to up to 13 polypeptides (mammals).

Subunit 2 (CO II) transfers the electrons from cytochrome c to the catalytic subunit 1. It contains two adjacent transmembrane regions in its N-terminus and the major part of the protein is exposed to the periplasmic or to the mitochondrial intermembrane space, respectively. CO II provides the substrate- binding site and contains a copper center called Cu(A), probably the primary acceptor in cytochrome c oxidase. An exception is the corresponding subunit of the cbb3-type oxidase which lacks the copper A redox-center. Several bacterial CO II have a C-terminal extension that contains a covalently bound heme c.

It has been shown [3,4] that nitrous oxide reductase (EC 1.7.99.6) (gene nosZ) of Pseudomonas has sequence similarity in its C-terminus to CO II. This enzyme is part of the bacterial respiratory system which is activated under anaerobic conditions in the presence of

nitrate or nitrous oxide. NosZ is a periplasmic homodimer that contains a dinuclear copper center, probably located in a 3- dimensional fold similar to the cupredoxin-like fold that has been suggested for the copper-binding site of CO II [3].

The dinuclear purple copper center is formed by 2 histidines and 2 cysteines [5]. This region was used as a signature pattern. The conserved valine and the conserved methionine are said to be involved in stabilizing the copper-binding fold by interacting with each other.

Consensus pattern V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M [The two C's and two H's are copper ligands] Sequences known to belong to this class detected by the patternALL, except for Paramecium primaurelia as well as in some plants where the pattern ends with Thr; an RNA editing event at this position could change this Thr to Met.

Note: cytochrome cbb(3) subunit 2 does not belong to this family.

[1] Capaldi R.A., Malatesta F., Darley-Usmar V.M. Biochim. Biophys. Acta 726:135-148(1983).

[2] Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B. J. Bacteriol. 176:5587-5600(1994).

[3] van der Oost J., Lappalainen P., Musacchio A., Warne A., Lemieux L., Rumbley J., Gennis R.B., Aasa R., Pascher T., Malmstrom B.G., Saraste M. EMBO J. 11:3209-3217(1992).

[4] Zumft W.G., Dreutsch A., Loechelt S., Cuypers H., Friedrich B., Schneider B. Eur. J. Biochem. 208:31-40(1992).

928. Cytochrome C assembly protein (CytC_asm)

This family consists of various proteins involved in cytochrome c assembly from mitochondria and bacteria; CycK from Rhizobium[3], CcmC from E. coli and Paracoccus denitrificans [2,1] and orf240 from wheat mitochondria [4]. The members of this family are probably integral membrane proteins with six predicted transmembrane helices. It has been proposed that members of this family comprise a membrane component of an ABC (ATP binding cassette) transporter complex. It is also proposed that this transporter is necessary for transport of some component needed for cytochrome c assembly. One member CycK

contains a putative heme-binding motif [3], orf240 also contains a putative heme-binding motif and is a proposed ABC transporter with c-type heme as its proposed substrate [4]. However it seems unlikely that all members of this family transport heme nor c-type apocytochromes because CcmC in the putative CcmABC transporter transports neither [1].

Number of members: 67

[1] Page D, Pearce DA, Norris HA, Ferguson SJ; Medline: 97195802 The Paracoccus denitrificans ccmA, B and C genes: cloning and sequencing, and analysis of the potential of their products to form a haem or apo-c-type cytochrome transporter. MICROBIOLOGY 1997;143:563-576.

[2] Thoeny-meyer L, Fischer F, Kunzler P, Ritz D, Hennecke H; Medline: 95362656 Escherichia coli genes required for cytochrome c maturation." J. BACTERIOL 1995;177:4321-4326.

[3] Delgado MJ, Yeoman KH, Wu G, Vargas C, Davies A, Poole RK, Johnston AWB, Downie JA; Medline: 95394794 Characterization of the cycHJKL genes involved in cytochrome c biogenesis and symbiotic nitrogen fixation in Rhizobium leguminosarum." J. BACTERIOL 1995;177:4927-4934.

[4] Bonnard G, Grienenberger JM; Medline: 95124303 A gene proposed to encode a transmembrane domain of an ABC transporter is expressed in wheat mitochondria." MOL. GEN. GENET 1995;246:91-99.

929. Cytochrome b559 subunits heme-binding site signature (cytochr_b559)

Cytochrome b559 [1] is an essential component of photosystem II complex from oxygenic photosynthetic organisms. It is an integral thylakoid membrane protein composed of two subunits, alpha (gene psbE) and beta (gene psbF), each of which contains a histidine residue located in a transmembrane region. The two histidines coordinate the heme iron of cytochrome b559.

The region around the heme-binding residue of both subunits is very similar and can be used as a signature pattern.

Consensus pattern[LIV]-x-[ST]-[LIVF]-R-[FYW]-x(2)-[IV]-H-[STGA]-[LIV]- [STGA]-[IV]-P [H is the heme iron ligand] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Pakrasi H.B., de Ciechi P., Whitmarsh J. EMBO J. 10:1619-1627(1991).

930. Cytochrome b/b6 signatures (Cytochrome_b)

In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex.

Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB and petD genes. The sequence of petB is colinear with the N-terminal part of mitochondrial cytochrome b, while petD corresponds to the C-terminal part. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups.

Apart from regions around some of the histidine heme ligands, there are a few conserved regions in the sequence of b/b6. The best conserved of these regions includes an invariant P-E-W triplet which lies in the loop that separates the fifth and sixth transmembrane segments. It seems to be important for electron transfer at the ubiquinone redox site - called Qz or Qo (where o stands for outside) - located on the outer side of the membrane.

A schematic representation of the structure of cytochrome b/b6 is shown below.

+---Fe-b562----+ | +---Fe-b566--|-+ | | | |

xxxxxxxxxxxHxHxxxxxxxxxxxxHxHxxxxxxxxxxPEWxxxxxxxxxxxxxxxxx <--------------

---Cytochrome-b---------------------------> <----Cytochrome-b6-petB---------><--Cytochrome-b6-petD----->

Two signature patterns were developed for cytochrome b/b6. The first includes the first conserved histidine of b/b6, which is a heme b562 ligand; the second includes the conserved PEW triplet.

Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the patternALL, except for 5 sequences.

Consensus pattern P-[DE]-W-[FY]-[LFY](2) Sequences known to belong to this class detected by the patternALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b.

[1] Howell N. J. Mol. Evol. 29:157-169(1989).
[2] Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993).

931. Phorbol esters / diacylglycerol binding domain (DAG_PE-bind)

Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) [1]. Phorbol esters can directly stimulate PKC. The N- terminal region of PKC, known as C1, has been shown [2] to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in the following proteins:

- Diacylglycerol kinase (EC 2.7.1.107) (DGK) [3], the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals.

- N-chimaerin. A brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown [4,5] to be able to bind phorbol esters.
- The raf/mil family of serine/threonine protein kinases. These protein kinases contain a single N-terminal copy of the DAG/PE-binding domain.
- The unc-13 protein from Caenorhabditis elegans. Its function is not known but it contains a copy of the DAG/PE-binding domain in its central section and has been shown to bind specifically to a phorbol ester in the presence of calcium [6].
- The vav oncogene. Vav was generated by a genetic rearrangement during gene transfer assays. Its expression seems to be restricted to cells of hematopoeitic origin. Vav seems [5,7] to contain a DAG/PE-binding domain in the central part of the protein.
- The Drosophila GTPase activating protein rotund.

The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. A signature pattern was developed that spans completely the DAG/PE domain.

Consensus pattern H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)- C-x(2)-C-x(4)-[HD]-x(2)-C-x(5,9)-C [All the C and H are involved in binding Zinc] Sequences known to belong to this class detected by the pattern ALL, except a few DGK's.

[1] Azzi A., Boscoboinik D., Hensey C. Eur. J. Biochem. 208:547-557(1992).
[2] Ono Y., Fujii T., Igarashi K., Kuno T., Tanaka C, Kikkawa U., Nishizuka Y. Proc. Natl. Acad. Sci. U.S.A. 86:4868-4871(1989).
[3] Sakane F., Yamada K., Kanoh H., Yokoyama C., Tanabe T. Nature 344:345-348(1990).
[4] Ahmed S., Kozma R., Monfries C., Hall C., Lim H.H., Smith P., Lim L. Biochem. J. 272:767-773(1990).
[5] Ahmed S., Kozma R., Lee J., Monfries C., Harden N., Lim L. Biochem. J. 280:233-241(1991).
[6] Ahmed S., Maruyama I.N., Kozma R., Lee J., Brenner S., Lim L. Biochem. J. 287:995-999(1992).
[7] Boguski M.S., Bairoch A., Attwood T.K., Michaels G.S. Nature 358:113-113(1992).

932. 3-dehydroquinate synthase (DHQ_synthase)

[1] Barten R, Meyer TF; Medline: 98273626 Cloning and characterisation of the Neisseria gonorrhoeae aroB gene." Mol Gen Genet 1998;258:34-44.
[2] Hawkins AR, Lamb HK; Medline: 96048023 The molecular biology of multidomain proteins. Selected examples." Eur J Biochem 1995;232:7-18.

The 3-dehydroquinate synthase EC:4.6.1.3 domain is present in isolation in various bacterial 3-dehydroquinate synthases and also present as a domain in the pentafunctional AROM polypeptide Swiss:P07547 [2]. 3-dehydroquinate (DHQ) synthase catalyses the formation of dehydroquinate (DHQ) and orthophosphate from 3-deoxy-D-arabino heptulosonic 7 phosphate [1]. This reaction is part of the shikimate pathway which is involved in the biosynthesis of aromatic amino acids.
Number of members: 25

933. Dihydrofolate reductase signature (DiHfolate_red)

Dihydrofolate reductases (EC 1.5.1.3) [1] are ubiquitous enzymes which catalyze the reduction of folic acid into tetrahydrofolic acid. They can be inhibited by a number of antagonists such as trimethroprim and methotrexate which are used as antibacterial or anticancerous agents. A signature pattern was derived from a region in the N-terminal part of these enzymes, which includes a conserved Pro-Trp dipeptide; the tryptophan has been shown [2] to be involved in the binding of substrate by the enzyme.

Consensus pattern[LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ] Sequences known to belong to this class detected by the patternALL, except for type II bacterial, plasmid-encoded, dihydrofolate reductases which do not belong to the same class of enzymes.

[1] Harpers' Review of Biochemistry, Lange, Los Altos (1985).
[2] Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C., Kraut J. J. Biol. Chem. 257:13650-13662(1982).

934. (DIL)

[1] Ponting CP; Medline: 95397417 AF-6/cno: neither a kinesin nor a myosin, but a bit of both." Trends Biochem Sci 1995;20:265-266.

Number of members: 31

935. (DNA_gyraseB_C)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

```
<----------------About-1400-residues----------------------->
[----------Protein 39-*-----][----Protein 52----] Phage T4
[----------gyrB-------*-----][--------gyrA----------------] Prokaryote II
 Archaebacteria
[----------parE-------*-----][--------parD----------------] Prokaryote IV
[--------------------*----------------------------------] Eukaryote and ASF
```

'*': Position of the pattern.

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the pattern ALL.

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).

[2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).

[3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).

[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

936. (DNA_topoisolIV)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes gyrA and gyrB [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes parC and parE). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

<----------------About-1400-residues----------------------->

[----------Protein 39-*-----][----Protein 52----] Phage T4

[----------gyrB-------*-----][--------gyrA-----------------] Prokaryote II Archaebacteria

[----------parE-------*-----][--------parD-----------------] Prokaryote IV

[---------------------*-----------------------------------] Eukaryote and ASF

'*': Position of the pattern.

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the patternALL.

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
[2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
[3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

937. Prolyl oligopeptidase family serine active site (DPPIV_N_term)

The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences.
- Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and argininyl residues.
- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.
- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.
- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).
- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

[1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
[2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
[3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).
[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

938. Deoxyhypusine synthase (DS)

Eukaryotic initiation factor 5A (eIF-5A) contains an unusual amino acid, hypusine [N epsilon-(4-aminobutyl-2-hydroxy)lysine]. The first step in the post-translational formation of hypusine is catalysed by the enzyme deoxyhypusine synthase (DS) EC:1.1.1.249. The modified version of eIF-5A, and DS, are required for eukaryotic cell proliferation [1].
Number of members: 9

[1] Liao DI, Wolff EC, Park MH, Davies DR; Medline: 98154315 Crystal structure of the NAD complex of human deoxyhypusine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site." Structure 1998;6:23-32.

939. (DUF21)

Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

940. (DUF59)

This family includes prokaryotic proteins of unknown function. The family also includes PhaH Swiss:O84984 from Pseudomonas putida. PhaH forms a complex with PhaF Swiss:O84982, PhaG Swiss:O84983 and PhaI Swiss:O84985, which hydroxylates phenylacetic acid to 2-hydroxyphenylacetic acid [1]. So members of this family may all be components of ring hydroxylating complexes.

Number of members: 15

[1] Olivera ER, Minambres B, Garcia B, Muniz C, Moreno MA, Ferrandez A, Diaz E, Garcia JL, Luengo JM; Medline: 98263372 Molecular characterization of the phenylacetic acid catabolic pathway in Pseudomonas putida U: the phenylacetyl-CoA catabolon." Proc Natl Acad Sci U S A 1998;95:6419-6424.

941. (DUF82)

The protein contains four conserved cysteines that may be involved in metal binding or disulphide bridges.

Number of members: 4

942. Riboflavin kinase / FAD synthetase (FAD_Synth)

This family consists part of the bifunctional enzyme riboflavin kinase / FAD synthetase. These enzymes have both ATP:riboflavin 5'-phospho transferase and ATP:FMN-adenylyltransferase activitys [1]. They catalyse the 5'-phosphorylation of riboflavin to FMN and the adenylylation of FMN to FAD [1].

CAUTION: It is not clear if this region of the enzymes catalyses either or both of the enzymatic reactions.

Number of members: 27

[1] Manstein DJ, Pai EF; Medline: 87057286 Purification and characterization of FAD synthetase from Brevibacterium ammoniagenes." J Biol Chem 1986;261:16169-16173.

943. [2Fe-2S] binding domain (fer2_2)

[1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber R; Medline: 96072968 Crystal structure of the xanthine oxidase-related aldehyde oxido-reductase from D. gigas." Science 1995;270:1170-1176.

Number of members: 53

944. Filovirus glycoprotein (Filo_glycop)

This family includes an extracellular region from the envelope glycoprotein of Ebola and Marburg viruses. This region is also produced as a separate transcript that gives rise to a non-structural, secreted glycoprotein, which is produced in large amounts and has an unknown function [1]. Processing of this protein may be involved in viral pathogenicity [2].

Number of members: 23

[1] Volchkov VE, Feldmann H, Volchkova VA, Klenk HD; Medline: 98245155 Processing of the Ebola virus glycoprotein by the proprotein convertase furin." Proc Natl Acad Sci U S A 1998;95:5762-5767.

[2] Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST; Medline: 96195018 The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing." Proc Natl Acad Sci U S A 1996;93:3602-3607.

945. Frataxin-like domain (Frataxin_Cyay)

This family contains proteins that have a domain related to the globular C-terminus of Frataxin the protein that is mutated in Friedreich's ataxia. This domain is found in a family of bacterial proteins. The function of this domain is currently unknown.

Number of members: 12

[1] Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P; Medline: 97084946 Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction." Trends Neurosci 1996;19:465-468.

946. (GAF)

Domain present in phytochromes and cGMP-specific phosphodiesterases.
Number of members: 296

[1] Aravind L, Ponting CP; Medline: 98094688 The GAF domain: an evolutionary link between diverse phototransducing proteins." Trends Biochem Sci 1997;22:458-459.

947. Galaptin signature (Gal-bind_lectin)

All vertebrates synthesize soluble galactoside-binding lectins [1,2,3] (also known as galectins, galaptins or S-lectin). These carbohydrate-binding proteins are developmentally regulated. Although their exact physiological role is not yet clear they seem to be involved in differentiation, cellular regulation and tissue construction. The sequence of galactoside-binding lectins from electric eel (electrolectin), conger eel (congerin), chicken and a number of mammalian species is known. These lectins are proteins of about 130 to 140 amino acid residues (14 Kd to 16 Kd).

A number of other proteins are known to belong to this family:
- Galectin-3 (also known as MAC-2 antigen; CBP-35 or IgE-binding protein), a 35 Kd lectin which binds immunoglobulin E and which is composed of two domains: a N-terminal domain that consist of tandem repeats of a glycine/ proline-rich sequence and a C-terminal galaptin domain.
- Galectin-4 [4], which is composed of two galaptin domains.
- Galectin-5.
- Galectin-7 [5], a keratinocyte protein which could be involved in cell-cell and/or cell-matrix interactions necessary for normal growth control.
- Galectin-8 [6], which is composed of two galaptin domains.

- Galectin-9 [7], which is composed of two galaptin domains.
- Human eosinophil lysophospholipase (EC 3.1.1.5) [8] (Charcot-Leyden crystal protein), a protein that may have both an enzymatic and a lectin activities. It forms hexagonal bipyramidal crystals in tissues and secretions from sites of eosinophil-associated inflammation.
- Caenorhabditis elegans 32 Kd lactose-binding lectin [9]. This lectin is composed of two galaptin domains.
- Caenorhabditis elegans lec-7 and lec-8.

One of the conserved regions of these lectins contains a tryptophan that has been shown [10] to be essential to the binding of galactosides. This region was used as a signature pattern for these proteins.

Consensus patternW-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQEGSKV]-x-[GH]-x(3)-[DENKHS]-[LIVMFC] [W binds carbohydrate] Sequences known to belong to this class detected by the pattern ALL, except for pig galectin 4.

[1] Barondes S.H., Gitt M.A., Leffler H., Cooper D.N.W. Biochimie 70:1627-1632(1988).
[2] Hirabayashi J., Kasai K.-I. J. Biochem. 104:1-4(1988).
[3] Barondes S.H., Castronovo V., Cooper D.N.W., Cummings R.D., Drickamer K., Feizi T., Gitt M.A., Hirabayashi J., Hughes C., Kasai K.-I., Leffler H., Liu F.-T., Lotan R., Mercurio A.M., Monsigny M., Pillair S., Poirer F., Raz A., Rigby P.W.J., Rini J.M., Wang J.L. Cell 76:597-598(1994).
[4] Oda Y., Herrmann J., Gitt M., Turck C.W., Burlingame A.L., Barondes S.H., Leffler H. J. Biol. Chem. 268:5929-5939(1993).
[5] Madsen P., Rasmussen H.H., Flint T., Gromov P., Kruse T.A., Honore B., Vorum H., Celis J.E. J. Biol. Chem. 270:5823-5829(1995).
[6] Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y. J. Biol. Chem. 270:3447-3453(1995).
[7] Wada J., Kanwar Y.S. J. Biol. Chem. 272:6078-6086(1997).
[8] Ackerman S.J., Corrette S.E., Rosenberg H.F., Bennett J.C., Mastrianni D.M., Nicholson-Weller A., Weller P.F., Chin D.T., Tenen D.G. J. Immunol. 150:456-468(1993).
[9] Hirabayashi J., Satoh M., Kasai K.-I. J. Biol. Chem. 267:15485-15490(1992).
[10] Abbott W.M., Feizi T. J. Biol. Chem. 266:5552-5557(1991).

948. (GARS) Phosphoribosylglycinamide synthetase signature (phosphoribosylamine glycine ligase)

PROSITE: PDOC00164; cross-reference(s): PS00184

[1] catalyzes the second step in the de novo biosynthesis of purine, the ATP-dependent addition of 5-phosphoribosylamine to glycine to form 5'phosphoribosylglycinamide.

In bacteria GARS is a monofunctional enzyme (encoded by the purD gene), in of a bifunctional enzyme (encoded by the ADE5,7 gene), in higher eukaryotes it is part, with AIRS and with phosphoribosylglycinamide formyltransferase (GART) of a trifunctional enzyme (GARS-AIRS-GART).

The sequence of GARS is well conserved. A highly conserved octapeptide was selected as a signature pattern.

Consensus patternR-F-G-D-P-E-x-[QM]

Sequences known to belong to this class detected by the patternALL.

[1]Aiba A., Mizobuchi K. J. Biol. Chem. 264:21239-21246(1989).

949. GLTT - GLTT repeat (12 copies)

This short repeat of unknown function is found in multiple copies in several C. elegans proteins. The repeat is five residues long and consists of XGLTT where X can be any amino acid. Number of members: 34.

950. Glu_synthase - Conserved region in glutamate synthase

This family represents a region of the glutamate synthase protein. This region is expressed as a seperate subunit in the glutamate synthase alpha subunit from archaebacteria, or part of a large multidomain enzyme in other organisms. The aligned region of these proteins contains a putative FMN binding site and Fe-S cluster. Number of members: 44.

[1] Medline: 97082505. Sequence of the GLT1 gene from Saccharomyces cerevisiae reveals the domain structure of yeast glutamate synthase. Filetici P, Martegani MP, Valenzuela L, Gonzalez A, Ballario P; Yeast 1996;12:1359-1366.

951. (Glyco_hydro_2) Glycosyl hydrolases family 2 signatures

GLYCOSYL_HYDROL_F2_1; PS00608; GLYCOSYL_HYDROL_F2_2

It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

-Beta-galactosidases (EC 3.2.1.23) from bacteria such as Escherichia coli (genes lacZ and ebgA), Clostridium acetobutylicum, Clostridium thermosulfurogenes, Klebsiella pneumoniae, Lactobacillus delbrueckii, or Streptococcus thermophilus and from the fungi Kluyveromyces lactis.

-Beta-glucuronidase (EC 3.2.1.31) from Escherichia coli (gene uidA) and from mammals.

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in Escherichia coli lacZ, to be the general acid/base catalyst in the active site of the enzyme. This region has been used as a signature pattern. A highly conserved region located some sixty residues upstream from the active site glutamate has been selected as a second signature pattern.

Consensus pattern N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4) Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [DENQLF]-[KRVW]-N-[HRY]-[STAPPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Rhizobium meliloti lacZ.

[1]Henrissat B. Biochem. J. 280:309-316(1991).

[2]Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. Gen. Microbiol. 137:369-380(1991).

[3]Gebler J.C., Aebersold R., Withers S.G. J. Biol. Chem. 267:11126-11130(1992).

952. (Glyco_hydro_3) Glycosyl hydrolases family 3 active site

PROSITE: PDOC00621. PROSITE cross-reference(s)PS00775; GLYCOSYL_HYDROL_F3

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

-Beta glucosidases (EC 3.2.1.21) from the fungi Aspergillus wentii (A-3), Hansenula anomala, Kluyveromyces fragilis, Saccharomycopsis fibuligera,(BGL1 and BGL2), Schizophyllum commune and Trichoderma reesei (BGL1).

-Beta glucosidases from the bacteria Agrobacterium tumefaciens (Cbg1), Butyrivibrio fibrisolvens (bglA), Clostridium thermocellum (bglB), Escherichia coli (bglX), Erwinia chrysanthemi (bgxA) and Ruminococcus albus.

-Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52).

-Bacillus subtilis hypothetical protein yzbA.

-Escherichica coli hypothetical protein ycfO and HI0959, the corresponding Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus wentii beta-glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern[LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]

Sequences known to belong to this class detected by the patternALL.

[1]Henrissat B. Biochem. J. 280:309-316(1991).

[2]Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).

[3]Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).

953. GP120 - Envelope glycoprotein GP120

The entry of HIV requires interaction of viral GP120 with Swiss:P01730 and a chemokine receptor on the cell surface. Number of members: 17891

[1]Medline: 98303379. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA; Nature 1998;393:648-659.

954. (GSPII_E) Bacterial type II secretion system protein E signature

PROSITE: PDOC00567. PROSITE cross-reference(s) PS00662; T2SP_E

A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1,2], have been found to be evolutionary related. These proteins are listed below:

-The 'E' protein from the GSP operon of: Aeromonas (gene exeE); Erwinia (gene outE); Escherichia coli (gene yheG); Klebsiella pneumoniae (gene pulE); Pseudomonas aeruginosa (gene xcpR); Vibrio cholerae (gene epsE) and Xanthomonas campestris (gene xpsE).

-Agrobacterium tumefaciens Ti plasmid virB operon protein 11. This protein is required for the transfer of T-DNA to plants.

-Bacillus subtilis comG operon protein 1 which is required for the uptake of DNA by competent Bacillus subtilis cells.

-Aeromonas hydrophila tapB, involved in type IV pilus assembly.

-Pseudomonas protein pilB, which is essential for the formation of the pili.

-Pseudomonas aeruginosa protein twitching mobility protein pilT.

-Neisseria gonorrhoeae type IV pilus assembly protein pilF.

-Vibrio cholerae protein tcpT, which is involved in the biosynthesis of the tcp pilus.

-Escherichia coli protein hofB (hopB).

-Escherichia coli hypothetical protein ygcB.

-Escherichia coli hypothetical protein yggR.

These proteins have from 344 (pilT and virB11) to 568 (tapB) amino acids, they are probably cytoplasmically located and, on the basis of the presence of a conserved P-loop region (see <PDOC00017>), probably bind ATP. A region that overlaps the 'B' motif of ATP-binding proteins was selected as a signature pattern.

Consensus pattern[LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D
Sequences known to belong to this class detected by the patternALL, except for ygcB.

[1]Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).
[2]Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).

955. (guanylate_cyc) Guanylate cyclases signature
PROSITE: PDOC00425. PROSITE cross-reference(s) PS00452;
GUANYLATE_CYCLASES Guanylate cyclases (EC 4.6.1.2) [1 to 4] catalyze the formation of cyclic GMP (cGMP) from GTP. cGMP acts as an intracellular messenger, activating cGMP dependent kinases and regulating CGMP-sensitive ion channels. The role of cGMP as a second messenger in vascular smooth muscle relaxation and retinal photo-transduction is well established. Guanylate cyclase is found both in the soluble and particular

fraction of eukaryotic cells. The soluble and plasma membrane-bound forms differ in structure, regulation and other properties.

Most currently known plasma membrane-bound forms are receptors for small polypeptides. The topology of such proteins is the following: they have a N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain, followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain that appears important for proper signalling and a cyclase catalytic domain. This topology is schematically represented below.

```
+----------------------xxxxx--------------------+--------------+
| Ligand-binding    XXXXX Protein Kinase like |  Cyclase    |
+----------------------xxxxx--------------------+--------------+
  Extracellular     Transmembrane      Cytoplasmic
```

The known guanylate cyclase receptors are:

-The sea-urchins receptors for speract and resact, which are small peptides that stimulate sperm motility and metabolism.

-The receptors for natriuretic peptides (ANF). Two forms of ANF receptors with guanylate cyclase activity are currently known: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP.

-The receptor for Escherichia coli heat-stable enterotoxin (GC-C). The endogenous ligand for this intestinal receptor seems to be a small peptide called guanylin.

-Retinal guanylate cyclase (retGC) which probably plays a specific functional role in the rods and/or cones of photoreceptors. It is not known if this protein acts as receptor, but its structure is similar to that of the other plasma membrane-bound GCs.

The soluble forms of guanylate cyclase are cytoplasmic heterodimers. The two subunits, alpha and beta are proteins of from 70 to 82 Kd which are highly related. Two forms of beta subunits are currently known: beta-1 which seems to be expressed in lung and brain, and beta-2 which is more abundant in kidney and liver.

The membrane and cytoplasmic forms of guanylate cyclase share a conserved domain which is probably important for the catalytic activity of the enzyme. Such a domain is also found twice in the different forms of membrane-bound adenylate cyclases (also known as

class-III) [5,6] from mammals, slime mold or Drosophila. A consensus pattern was derived from the most conserved region in that domain.

Consensus patternG-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-[DNTA]-x(5)-[DE]

Sequences known to belong to this class detected by the patternALL, except for the sea urchin Arbacia punctulata resact receptor which lack this domain.

Note this pattern will detect both domains of adenylate cyclases class-III.

[1]Koesling D., Boehme E., Schultz G. FASEB J. 5:2785-2791(1991).

[2]Garbers D.L. New Biol. 2:499-504(1990).

[3]Garbers D.L. Cell 71:1-4(1992).

[4]Yuen P.S.T., Garbers D.L. Annu. Rev. Neurosci. 15:193-225(1992).

[5]Iyengar R. FASEB J. 7:768-775(1993).

[6]Barzu O., Danchin A. Prog. Nucleic Acid Res. Mol. Biol. 49:241-283(1994).

956. Hemolysin-type calcium-binding region signature (HemolysinCabinD)

Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which are currently known to belong to this category are:

- Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: E. coli (gene hlyA), A. pleuropneumoniae (gene appA), A. actinomycetemcomitans and P. haemolytica (leukotoxin) (gene lktA).
- Cyclolysin from Bordetella pertussis (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin.
- Extracellular zinc proteases: serralysin (EC 3.4.24.40) from Serratia, prtB and prtC from Erwinia chrysanthemi and aprA from Pseudomonas aeruginosa.
- Nodulation protein nodO from Rhizobium leguminosarum.

A signature pattern was derived from conserved positions in the sequence of the calcium-binding domain.

Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D Sequences known to belong to this class detected by the pattern ALL.

Note: This pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins.

[1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).

[2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).

957. Hint module (Hint)

This is an alignment of the Hint module in the Hedgehog proteins. It does not include any Inteins which also possess the Hint module.

Number of members: 36

[1] Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ; Medline: 97474313 Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins." Cell 1997;91:85-97.

958. Hydantoinase/oxoprolinase (Hydantoinase)

This family includes the enzymes hydantoinase and oxoprolinase EC:3.5.2.9. Both reactions involve the hydrolysis of 5-membered rings via hydrolysis of their internal imide bonds [1].

Number of members: 14

[1] Ye GJ, Breslow EB, Meister A, Guo-jie GE$[corrected to Ye GJ]; Medline: 97113037 The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA cloning" [published erratum appears in J Biol Chem 1997 Feb 14;272(7):4646] J Biol Chem 1996;271:32293-32300.

959. IMP dehydrogenase / GMP reductase signature (IMPDH_N)

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].

GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides.

IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

Consensus pattern[LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).
[2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).
[3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).

960. impB/mucB/samB family (IMS)

These proteins are involved in UV protection (Swiss).
Number of members: 38

961. Type II intron maturase (Intron_maturas2)

Group II introns use intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA [2]. Although this type of intron is self splicing in vitro they require a maturase protein for splicing in vivo. It has been shown that a specific region of the aI2 intron is needed for the maturase function [1]. This region was found to be conserved in group II introns and called domain X [3].

Number of members: 335

[1] Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P; Medline: 94301788 Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. Nucleic Acids Res 1994;22:2057-2064.

[2] Guo H, Zimmerly S, Perlman PS, Lambowitz AM; Medline: 98031910 Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA." EMBO J 1997;16:6835-6848.

[3] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

962. LAGLIDADG endonuclease (Intron_maturase)

[1] Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL; Medline: 97331323 The structure of I-CreI, a group I intron-encoded homing endonuclease." Nat Struct Biol 1997;4:468-476.

[2] Belfort M, Roberts RJ; Medline: 97402526 Homing endonucleases: keeping the house in order." Nucleic Acids Res 1997;25:3379-3388.

[3] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

Number of members: 220

963. Isopentenyl transferase (IPT)

Isopentenyl transferase / dimethylallyl transferase synthesizes isopentenyladensosine 5'-monophosphate, a cytokinin that induces shoot formation on host plants infected with the Ti plasmid [1].

Number of members: 16

[1] Canaday J, Gerad JC, Crouzet P, Otten L; Medline: 93101133 "Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4." Mol Gen Genet 1992;235:292-303.

964. Laminin EGF-like (Domains III and V) (laminin_EGF)

This family is like EGF but has 8 conserved cysteines instead of 6.

Number of members: 501

[1] Engel J; Medline: 93041759 Laminins and other strange proteins." Biochemistry 1992;31:10643-10651.

965. Legume lectins signatures (lectin_legA)

Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals).

Legume lectins are synthesized as precursor proteins of about 230 to 260 amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal). The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain.

Two signature patterns were developed specific to legume lectins: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for

the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain.

Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST] Sequences known to belong to this class detected by the pattern ALL.

[1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).
[2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).

966. Malate synthase signature (malate_synthase)

Malate synthase (EC 4.1.3.2) catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form malate - the second step of the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. Malate synthase is a protein of 530 to 570 amino acids whose sequence is highly conserved across species [1]. As a signature pattern, a very conserved region was selected in the central section of the enzyme.

Consensus pattern[KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F Sequences known to belong to this class detected by the pattern ALL.

[1] Bruinenberg P.G., Blaauw M., Kazemier B., Ab G. Yeast 6:245-254(1990).

967. MatK/TrnK amino terminal region (MatK_N)

[1] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

Number of members: 495

968. MOZ/SAS family (MOZ_SAS)

This region of these proteins has been suggested to be homologous to acetyltransferases [1]. However the similarity is not supported by standard sequence analysis.

Number of members: 15

[1] Kamine J, Elangovan B, Subramanian T, Coleman D, Chinnadurai G; Medline: 96182937 Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator." Virology 1996;216:357-366.

[2] Reifsnyder C, Lowell J, Clarke A, Pillus L; Medline: 96376969 Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases" [see comments] [published erratum appears in Nat Genet 1997 May;16(1):109] Nat Genet 1996;14:42-49.

969. mRNA capping enzyme (mRNA_cap_enzyme)

[1] Hakansson K, Doherty AJ, Shuman S, Wigley DB; Medline: 97304383 X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes." Cell 1997;89:545-553.

Number of members: 7

970. DNA mismatch repair proteins mutS family signature (MutS_C)

Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related [2,3]. One of these families is called mutS [4,E1], it consists of:

- Prokaroytic protein mutS protein (also called hexA in Streptococcus pneumoniae). Muts is thought to carry out the mismatch recognition step of DNA repair.
- Eukaryotic MSH1, which is involved in mitochondrial DNA repair.
- Eukaryotic MSH2, which is involved in nuclear postreplication mismatch repair. MSH2 heterodimerizes with MSH6. In man, MSH2 is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).

- Eukaryotic MSH3, which is probably involved in the repair of large loops.
- Eukaryotic MSH4, which is involved in meiotic recombination.
- Eukaryotic MSH5, which is involved in meiotic recombination.
- Eukaryotic MSH6 (also known as G/T mismatch binding protein), a DNA-repair protein that binds to G/T mismatches through heterodimerization with MSH2.
- Prokaryotic protein mutS2 whose function is not yet known.
- A coral (Sarcophyton glaucum) mitochondrial encoded mutS-like protein.

As a signature pattern for this class of mismatch repair proteins a region rich in glycine and negatively charged residues was selected This region is found in the C-terminal section of these proteins; about 80 residues to the C-terminal of an ATP-binding site motif 'A' (P-loop) (see <PDOC00017>).

Consensus pattern[ST]-[LIVMF]-x-[LIVM]-x-D-E-[LIVMFY]-[GC]-[RKH]-G-[GST]- x(4)-G Sequences known to belong to this class detected by the pattern ALL, except for mutS2.

[1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987).
[2] Haber L.T., Walker G.C. EMBO J. 10:2707-2715(1991).
[3] New L., Liu K., Crouse G.F. Mol. Gen. Genet. 239:97-108(1993).
[4] Eisen J.A. Nucleic Acids Res. 26:4291-4300(1998).

971. MutS family, N-terminal putative DNA binding domain (MutS_N)

This family consists of the N-terminal region of proteins in the mutS family of DNA mismatch repair proteins and is found associated with MutS_C located in the C-terminal region. The mutS family of proteins is named after the salmonella typhimurium MutS protein involved in mismatch repair; other members of the family included the eukaryotic MSH 1,2,3,4,5 and 6 proteins. These have various roles in DNA repair and recombination. Human MSH has been implicated in non-polyposis colorectal carcinoma (HNPCC) and is a mismatch binding protein [2]. The aligned region corresponds in part with domains A1, A2 (which may bind DNA) and B (which binds dsDNA in vitro) from T. thermophilus MutS as characterised in [1].

Number of members: 43

972. Domain in Myosin and Kinesin Tails (MyTH4)

Domain present twice in myosin-VIIa, and also present in 3 other myosins.

[1] Chen ZY, Hasson T, Kelley PM, Schwender BJ, Schwartz MF, Ramakrishnan M, Kimberling WJ, Mooseker MS, Corey DP; Medline: 97038686 Molecular cloning and domain structure of human myosin-VIIa, the gene product defective in Usher syndrome 1B." Genomics 1996;36:440-448.

Number of members: 21

973. Sodium and potassium ATPases beta subunits signatures (Na_K-ATPase)

The sodium pump (Na+,K+ ATPase), located in the plasma membrane of all animal cells [1], is an heterotrimer of a catalytic subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta chain) and a small hydrophobic protein of about 6 Kd. The beta subunit seems [2] to regulate, through the assembly of alpha/beta heterodimers, the number of sodium pumps transported to the plasma membrane.

Structurally the beta subunit is composed of a charged cytoplasmic domain of about 35 residues, followed by a transmembrane region, and a large extracellular domain that contains three disulfide bonds and glycosylation sites. This structure is schematically represented in the figure below.

```
+----+ +--+ +-----------+||||||
xxxxxxxxxxxxxxxxxxxxxxxxCxxxxCxCxxCxxxxxxxCxxxxxxxxxxxCxxxxx
**** **** <-Cyt-><TM><------------Extracellular-------------------->
```

'C': conserved cysteine involved in a disulfide bond.
'*': position of the patterns.

Two isoforms of the beta subunit (beta-1 and beta-2) are currently known; they share about 50% sequence identity. Gastric (K+, H+) ATPase (proton pump) responsible for acid production in the stomach consist of two subunits [3]; the beta chain is highly similar to the sodium pump beta subunits. Two signature patterns were developed for beta subunits. The

first is located in the cytoplasmic domain, while the second is found in the extracellular domain and contains two of the cysteines involved in disulfide bonds.

Consensus pattern [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G [The two C's are involved in disulfide bonds] Sequences known to belong to this class detected by the patternALL, except for the beta subunit of the sodium pump of brine shrimp whose sequence is highly divergent in that region.

[1] Horisberger J.D., Lemas V., Krahenbul J.P., Rossier B.C. Annu. Rev. Physiol. 53:565-584(1991).

[2] McDonough A.A., Gerring K., Farley R.A. FASEB J. 4:1598-1605(1990).

[3] Toh B.-H., Gleeson P.A., Simpson R.J., Moritz R.L., Callaghan J.M., Goldkorn I., Jones C.M., Martinelli T.M., Mu F.-T., Humphris D.C., Pettitt J.M., Mori Y., Masuda T., Sobieszczuk P., Weinstock J., Mantamadiotis T., Baldwin G.S. Proc. Natl. Acad. Sci. U.S.A. 87:6418-6422(1990).

974. Respiratory-chain NADH dehydrogenase subunit 1 signatures (NADHdh)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there are fifteen which are located in the membrane part, seven of which are encoded by the mitochondrial and chloroplast genomes of most species. The most conserved of these organelle-encoded subunits is known as subunit 1 (gene ND1 in mitochondrion, and NDH1 in chloroplast) and seems to contain the ubiquinone binding site.

The ND1 subunit is highly similar to subunit 4 of Escherichia coli formate hydrogenlyase (gene hycD), subunit C of hydrogenase-4 (gene hyfC). Paracoccus denitrificans NQO8 and Escherichia coli nuoH NADH-ubiquinone oxidoreductase subunits also belong to this family [3]. Two signature patterns were developed based on conserved regions of this subunit.

Consensus pattern G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]-[LIVMFTA]- K-[LVMYST]-[LIVMFYG]-x-[KR]-[EQG] Sequences known to belong to this class detected by the patternALL, except for watermelon and Leishmania ND1.

Consensus pattern P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G Sequences known to belong to this class detected by the pattern ALL, except for Chlamydomonas reinhardtii and Pisaster ochraceus ND1, and tobacco NDH1.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
[3] Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

975. Nickel-dependent hydrogenases large subunit signatures (NiFeSe_Hases)

Hydrogenases are enzymes that catalyze the reversible activation of hydrogen and which occur widely in prokaryotes as well as in some eukaryotes. There are various types of hydrogenases, but all of them seem to contain at least one iron-sulfur cluster. They can be broadly divided into two groups: hydrogenases containing nickel and, in some cases, also selenium (the [NiFe] and [NiFeSe] hydrogenases) and those lacking nickel (the [Fe] hydrogenases).

The [NiFe] and [NiFeSe] hydrogenases are heterodimer that consist of a small subunit that contains a signal peptide and a large subunit. All the known large subunits seem to be evolutionary related [1]; they contain two Cys-x-x- Cys motifs; one at their N-terminal end; the other at their C-terminal end. These four cysteines are involved in the binding of nickel [2]. In the [NiFeSe] hydrogenases the first cysteine of the C-terminal motif is a selenocysteine which has experimentally been shown to be a nickel ligand [3]. Two patterns were developed which are centered on the Cys-x-x-Cys motifs.

Alcaligenes eutrophus possess a NAD-reducing cytoplasmic hydrogenase (hoxS) [4]; this enzyme is composed of four subunits. Two of these subunits (beta and delta) are responsible for the hydrogenase reaction and are evolutionary related to the large and small subunits of

membrane-bound hydrogenases. The alpha subunit of coenzyme F420 hydrogenase (EC 1.12.99.1) (FRH) from archaebacterial methanogens also belongs to this family.

Consensus pattern R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

[1] Menon N.K., Robbins J., Peck H.D. Jr., Chatelus C.Y., Choi E.-S., Przybyla A.E. J. Bacteriol. 172:1969-1977(1990).
[2] Volbeda A., Charon M.-H., Piras C., Hatchikian E.C., Frey M., Fontecilla-Camps J.C. Nature 373:580-587(1995).
[3] Eidsness M.K., Scott R.A., Prickrill B., der Vartaninan D.V., LeGall J., Moura I., Moura J.J.G., Peck H.D. Jr. Proc. Natl. Acad. Sci. U.S.A. 86:147-151(1989).
[4] Tran-Betcke A., Warnecke U., Boecker C., Zaborosch C., Friedrich B. J. Bacteriol. 172:2920-2929(1990).

976. NADH-Ubiquinone oxidoreductase (complex I), chain 5 C-terminus (oxidored_q1_C)

This sub-family represents a carboxyl terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 from chloroplasts are in this family. This sub-family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.

Number of members: 572

[1] Walker JE; Medline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.

977. NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus (oxidored_q1_N)

This sub-family represents an amino terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 and eubacterial chain L are in this family. This sub-family is part of

complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.

Number of members: 546

[1] Walker JE; Medline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.

978. oxidored_q2. NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4L (EC 1.6.5.3). ND4L OR NAD4L. Arabidopsis thaliana (Mouse-ear cress). Mitochondrion. OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.

[1] SEQUENCE FROM N.A. MEDLINE; 93156682. Brandt P., Sunkel S., Unseld M., Brennicke A., Knoop V.; "The nad4L gene is encoded between exon c of nad5 and orf25 in the Arabidopsis mitochondrial genome."; Mol. Gen. Genet. 236:33-38(1992).

[2] SEQUENCE FROM N.A. STRAIN=CV. COLUMBIA; MEDLINE; 97141919 Unseld M., Marienfeld J.R., Brandt P., Brennicke A.; "The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides."; Nat. Genet. 15:57-61(1997).

979. oxidored_q4. Protein name NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 3, CHLOROPLAST. Synonym(s)EC 1.6.5.3. Gene name(s)NDHC OR NDH3 From Zea mays (Maize) Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.

SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.

[1] SEQUENCE FROM N.A. MEDLINE; 89281491. Steinmueller K., Ley A.C., Steinmetz A.A., Sayre R.T., Bogorad L.; "Characterization of the ndhC-psbG-ORF157/159 operon of maize plastid DNA and of the cyanobacterium Synechocystis sp. PCC6803."; Mol. Gen. Genet. 216:60-69(1989).

[2] SEQUENCE FROM N.A. MEDLINE; 95395841. Maier R.M., Neckermann K., Igloi G.L., Koessel H.; "Complete sequence of the maize chloroplast genome: gene content,

hotspots of divergence and fine tuning of genetic information by transcript editing."; J. Mol. Biol. 251:614-628(1995).

980. PAC: PAC motif

PAC motif occurs C-terminal to a subset of all known PAS motifs. It is proposed to contribute to the PAS domain fold [3]. Number of members: 181

[1] Medline: 97446881 PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. Zhulin IB, Taylor BL, Dixon R; Trends Biochem Sci 1997;22:331-333.

[2] Medline: 95275818. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. Borgstahl GE, Williams DR, Getzoff ED; Biochemistry 1995;34:6278-6287.

[3] Medline: 98044337. PAS: a multifunctional domain family comes to light. Ponting CP, Aravind L; Curr Biol 1997;7:674-677.

981. PARP: Poly(ADP-ribose) polymerase catalytic region.

Poly(ADP-ribose) polymerase catalyses the covalent attachment of ADP-ribose units from NAD+ to itself and to a limited number of other DNA binding proteins, which decreases their affinity for DNA. Poly(ADP-ribose) polymerase is a regulatory component induced by DNA damage.

The carboxyl-terminal region is the most highly conserved region of the protein. Experiments have shown that a carboxyl 40 kDa fragment is still catalytically active [2]. Number of members: 19

[1] Medline: 96353841 Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE; Proc Natl Acad Sci U S A 1996;93:7481-7485.

[2] Medline: 93293867 The carboxyl-terminal domain of human poly(ADP-ribose) polymerase. Overproduction in Escherichia coli, large scale purification, and characterization. Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G, Althaus FR; J Biol Chem 1993;268:13454-13461.

982. PC_rep: Proteasome/cyclosome repeat

[1] Medline: 97348748 A repetitive sequence in subunits of the 26S proteasome and 20S cyclosome (anaphase-promoting complex). Lupas A, Baumeister W, Hofmann K; Trends Biochem Sci 1997;22:195-196.

Number of members: 112

983. Peptidase_M1: Peptidase family M1

Members of this family are aminopeptidases. The members differ widely in specificity, hydrolysing acidic, basic or neutral N-terminal residues. This family includes leukotriene-A4 hydrolase Swiss:P09960, this enzyme also has an aminopeptidase activity [1]. Number of members: 72

[1] Medline: 95405261 Evolutionary families of metallopeptidases. Rawlings ND, Barrett AJ; Meth Enzymol 1995;248:183-228.

984. Neutral zinc metallopeptidases, zinc-binding region signature (Peptidase_M8)

PROSITE cross-reference(s) PS00142; ZINC_PROTEASE

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily,known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

Family M1

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
- Mammalian aminopeptidase N (EC 3.4.11.2).
- Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
- Yeast aminopeptidase yscII (gene APE2).
- Yeast alanine/arginine aminopeptidase (gene AAP1).
- Yeast hypothetical protein YIL137c.
- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

Family M2

- Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.

Family M3

- Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.
- Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).
- Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
- Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).
- Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).
- Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).
- Yeast hypothetical protein YKL134c.

Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus.
- Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB).
- Extracellular elastase from Staphylococcus epidermidis.
- Extracellular protease prt1 from Erwinia carotovora.
- Extracellular minor protease smp from Serratia marcescens.
- Vibriolysin (EC 3.4.24.25) from various species of Vibrio.
- Protease prtA from Listeria monocytogenes.
- Extracellular proteinase proA from Legionella pneumophila.

Family M5

- Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.

Family M6

- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

Family M7

- Streptomyces extracellular small neutral proteases

Family M8

- Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of Leishmania.

Family M9

- Microbial collagenase (EC 3.4.24.3) from Clostridium perfringens and Vibrio alginolyticus.

Family M10A

- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from Serratia.
- Alkaline metalloproteinase from Pseudomonas aeruginosa (gene aprA).
- Secreted proteases A, B, C and G from Erwinia chrysanthemi.
- Yeast hypothetical protein YIL108w.

Family M10B

- Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).
- Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A proteas that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.
- Soybean metalloendoproteinase 1.

Family M11

- Chlamydomonas reinhardtii gamete lytic enzyme (GLE).

Family M12A

- Astacin (EC 3.4.24.21), a crayfish endoprotease.
- Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border

metalloendopeptidase.

- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The Drosophila homolog of BMP-1 is the dorsal-ventral patterning protein tolloid.

- Blastula protease 10 (BP10) from Paracentrotus lividus and the related protein SpAN from Strongylocentrotus purpuratus.

- Caenorhabditis elegans protein toh-2.

- Caenorhabditis elegans hypothetical protein F42A10.8.

- Choriolysins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish Oryzias lapides. These proteases participates in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.

Family M12B

- Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53).

- Mouse cell surface antigen MS2.

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).

- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.

- Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.

- Peptidase O from Lactococcus lactis (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

Family M30

- Staphylococcus hyicus neutral metalloprotease.

Family M32

- Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from Thermus aquaticus which is most active at high temperature.

Family M34

- Lethal factor (LF) from Bacillus anthracis, one of the three proteins composing the anthrax toxin.

Family M35

- Deuterolysin (EC 3.4.24.39) from Penicillium citrinum and related proteases from various species of Aspergillus.

Family M36

- Extracellular elastinolytic metalloproteinases from Aspergillus.

From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ]

[The two H's are zinc ligands] [E is the active site residue]

Sequences known to belong to this class detected by the patternALL, except for members of families M5, M7 amd M11.

Other sequence(s) detected in SWISS-PROT57; including Neurospora crassa conidiation-specific protein 13 which could be a zinc-protease.

[1]Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989).

[2]Murphy G.J.P., Murphy G., Reynolds J.J. FEBS Lett. 289:4-7(1991).
[3]Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B., Stoecker W. Zoology 99:237-246(1996).
[4]Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).
[5]Woessner J. Jr. FASEB J. 5:2145-2154(1991).
[6]Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992).
[7]Montecucco C., Schiavo G. Trends Biochem. Sci. 18:324-327(1993).
[8]Niemann H., Blasi J., Jahn R. Trends Cell Biol. 4:179-185(1994).

985. PHO4: Phosphate transporter family

This family includes PHO-4 from Neurospora crassa which is a is a Na(+)-phosphate symporter [1]. This family also contains the leukemia virus receptor Swiss:Q08344. Number of members: 41

[1] Medline: 95249577 Repressible cation-phosphate symporters in Neurospora crassa. Versaw WK, Metzenberg RL; Proc Natl Acad Sci U S A 1995;92:3884-3887.

986. Photosynthetic reaction center proteins signature (photoRC)

PROSITE cross-reference(s): PS00244; REACTION_CENTER

In the photosynthetic reaction center of purple bacteria, two homologous integral membrane proteins, L(ight) and M(edium), are known to be essential to the light-mediated water-splitting process. In the photosystem II of eukaryotic chloroplasts two related proteins are involved: the D1 (psbA) and D2 proteins (psbD). These four types of protein probably evolved from a common ancestor [see 1,2 for recent reviews].

A signature pattern was developed which include two conserved histidine residues. In L and M chains, the first histidine is a ligand of the magnesium ion of the special pair bacteriochlorophyll, the second is a ligand of a ferrous non-heme iron atom. In photosystem II these two histidines are thought to play a similar role.

Consensus pattern[NQH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2)
[The first H is a magnesium ligand] [The second H is a iron ligand]
Sequences known to belong to this class detected by the patternALL, except

for broad bean psbA which has Gln instead of the second His.

[1]Michel H., Deisenhofer J. Biochemistry 27:1-7(1988).
[2]Barber J. Trends Biochem. Sci. 12:321-326(1987).

987. phytochrome: Phytochrome region

This family contains a region specific to phytochrome proteins. Number of members: 145

988. PI3K_C2: C2 domain

Phosphoinositide 3-kinase region postulated to contain a C2 domain. Outlier of C2 family. Number of members: 39

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.
[2] Medline: 97398940 Phosphoinositide 3-kinases: a conserved family of signal transducers. Vanhaesebroeck B, Leevers SJ, Panayotou G, Waterfield MD; Trends Biochem Sci 1997;22:267-272.

989. PI3Ka: Phosphoinositide 3-kinase family, accessory domain (PIK domain)

PIK domain is conserved in all PI3 and PI4-kinases. Its role is unclear but it has been suggested [2] to be involved in substrate presentation.
Number of members: 47

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.
[2] Medline: 94069320 Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability. Flanagan CA, Schnieders EA, Emerick AW, Kunisawa R, Admon A, Thorner J; Science 1993;262:1444-1448.

990. P-II protein signatures

PROSITE cross-reference(s): PS00496; PII_GLNB_UMP, PS00638; PII_GLNB_CTER

The P-II protein (gene glnB) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme. Conversely, in nitrogen excess, P-II-UMP is deuridylated and then promotes the adenylation of GS. P-II also indirectly controls the transcription of the GS gene (glnA) by preventing NR-II (ntrB) to phosphorylate NR-I (ntrC) which is the transcriptional activator of glnA. Once P-II is uridylylated, these events are reversed.

P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is urydylated is located in the central part of the protein.

In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being urydylated.

In methanogenic archaebacteria, the nitrogenase iron protein gene (nifH) is followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation.

In the red alga, Porphyra purpurea, there is a glnB homolog encoded in the chloroplast genome.

Other proteins highly similar to glnB are:

- Bacillus subtilis protein nrgB [5].
- Escherichia coli hypothetical protein ybaI [6].

Two signature patterns were developed for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the urydylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein.

Consensus patternY-[KR]-G-[AS]-[AE]-Y [The second Y is uridylated]
Sequences known to belong to this class detected by the patternALL glnB's from eubacteria.

Consensus pattern[ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM]

[1]Magasanik B. Biochimie 71:1005-1012(1989).
[2]Holtel A., Merrick M. Mol. Gen. Genet. 215:134-138(1988).
[3]Cheah E., Carr P.D., Suffolk P.M., Vasuvedan S.G., Dixon N.E., Ollis D.L. Structure 2:981-990(1994).
[4]Sibold L., Henriquet M., Possot O., Aubert J.-P. Res. Microbiol. 142:5-12(1991).
[5]Wray L.V. Jr., Atkinson M.R., Fisher S.H. J. Bacteriol. 176:108-114(1994).
[6]Allikmets R., Gerrard B.C., Court D., Dean M.C. Gene 136:231-236(1993).

991. PIP5K: Phosphatidylinositol-4-phosphate 5-Kinase

This family contains a region from the common kinase core found in the type I phosphatidylinositol-4-phosphate 5-kinase (PIP5K) family as described in [1]. The family consists of various type I, II and III PIP5K enzymes. PIP5K catalyses the formation of phosphoinositol-4,5-bisphosphate via the phosphorylation of phosphatidylinositol-4-phosphate a precursor in the phosphinositide signaling pathway. Number of members: 33

[1] Medline: 98204859. Type I phosphatidylinositol-4-phosphate 5-kinases. Cloning of the third isoform and deletion/substitution analysis of members of this novel lipid kinase family. Ishihara H, Shibasaki Y, Kizuki N, Wada T, Yazaki Y, Asano T, Oka Y; J Biol Chem 1998;273:8741-8748.
[2] Medline: 97115834 Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family. Loijens JC, Anderson RA; J Biol Chem 1996 20;271:32937-32943.

992. PolyA_pol: Poly A polymerase family

This family includes nucleic acid independent RNA polymerases, such as Poly(A) polymerase, which adds the poly (A) tail to mRNA EC:2.7.7.19. This family also includes the tRNA nucleotidyltransferase that adds the CCA to the 3' of the tRNA EC:2.7.7.25. Number of members: 31

[1] Medline: 93066242 Identification of the gene for an Escherichia coli poly(A) polymerase. Cao GJ, Sarkar N; Proc Natl Acad Sci U S A 1992;89:10380-10384.

993. Photosystem I psaA and psaB proteins signature (psaA_psaB)
PROSITE cross-reference(s)PS00419; PHOTOSYSTEM_I_PSAAB

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. PSI is found in the chloroplast of plants and cyanobacteria. The electron transfer components of the reaction center of PSI are a primary electron donor P-700 (chlorophyll dimer) and five electron acceptors: A0 (chlorophyll), A1 (a phylloquinone) and three 4Fe-4S iron-sulfur centers: Fx, Fa, and Fb.

PsaA and psaB, two closely related proteins, are involved in the binding of P700, A0, A1, and Fx. psaA and psaB are both integral membrane proteins of 730 to 750 amino acids that seem to contain 11 transmembrane segments. The Fx 4Fe-4S iron-sulfur center is bound by four cysteines; two of these cysteines are provided by the psaA protein and the two others by psaB. The two cysteines in both proteins are proximal and located in a loop between the ninth and tenth transmembrane segments. A leucine zipper motif seems to be present [2] downstream of the cysteines and could contribute to dimerization of psaA/psaB.

The signature pattern for these proteins is based on the perfectly conserved region that includes the two iron-sulfur binding cysteines.
Consensus patternC-D-G-P-G-R-G-G-T-C [The two C's bind the iron-sulfur center]

[1]Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).
[2]Webber A.N., Malkin R. FEBS Lett. 264:1-14(1990).

994. PSBH: Photosystem II 10 kDa phosphoprotein
This protein is phosphorylated in a light dependent reaction.
Number of members: 20

995. PsbJ
This family consists of the photosystem II reaction center protein PsbJ from plants and Cyanobacteria. In Synechocystis sp. PCC 6803 PsbJ regulates the number of photosystem II centers in thylakoid membranes, it is a predicted 4kDa protein with one membrane spanning domain [1]. Number of members: 20

[1] Medline: 93131892. Genetic and immunological analyses of the cyanobacterium Synechocystis sp. PCC 6803 show that the protein encoded by the psbJ gene regulates the number of photosystem II centers in thylakoid membranes. Lind LK, Shukla VK, Nyhus KJ, Pakrasi HB; J Biol Chem 1993;268:1575-1579.

996. PSBT: Photosystem II reaction centre T protein
The exact function of this protein is unknown. It probably consists of a single transmembrane spanning helix. The Swiss:P37256 protein, appears to be (i) a novel photosystem II subunit and (ii) required for maintaining optimal photosystem II activity under adverse growth conditions [1]. Number of members: 17

[1] Medline: 94298765. The chloroplast ycf8 open reading frame encodes a photosystem II polypeptide which maintains photosynthetic activity under adverse growth conditions. Monod C, Takahashi Y, Goldschmidt-Clermont M, Rochaix JD; EMBO J 1994;13:2747-2754.

997. PSI_8. PHOTOSYSTEM I REACTION CENTRE SUBUNIT VIII. Synonym(s)PSI-I. Gene name(s)PSAI. From Hordeum vulgare (Barley). Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
MAY HELP IN THE ORGANIZATION OF THE PSAL SUBUNIT. BELONGS TO THE PSAI FAMILY.

[1] SEQUENCE FROM N.A. MEDLINE; 90036933. Scheller H.V., Okkels J.S., Hoej P.B., Svendsen I., Roepstorff P., Moeller B.L.; "The primary structure of a 4.0-kDa photosystem I polypeptide encoded by the chloroplast psaI gene."; J. Biol. Chem. 264:18402-18406(1989).

998. PSI_PsaJ: Photosystem I reaction centre subunit IX / PsaJ
This family consists of the photosystem I reaction centre subunit IX or PsaJ from various organisms including Synechocystis sp. (strain pcc 6803), Pinus thunbergii (green pine) and Zea mays (maize). PsaJ Swiss:P19443 is a small 4.4kDa, chloroplastal encoded, hydrophobic subunit of the photosystem I reaction complex its function is not yet fully understood [1]. PsaJ can be cross-linked to PsaF Swiss:P12356 and has a single predicted transmembrane

domain it has a proposed role in maintaing PsaF in the correct orientation to allow for fast electron transfer from soluble donor proteins to P700+ [1]. Number of members: 18

[1] Medline: 99238330. A large fraction of PsaF is nonfunctional in photosystem I complexes lacking the PsaJ subunit. Fischer N, Boudreau E, Hippler M, Drepper F, Haehnel W, Rochaix JD; Biochemistry 1999;38:5546-5552.
[2] Medline: 93252282. Genes encoding eleven subunits of photosystem I from the thermophilic cyanobacterium Synechococcus sp. Muhlenhoff U, Haehnel W, Witt H, Herrmann RG; Gene 1993;127:71-78.

999. PSII. Protein namePHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN. Synonym(s)CP-47 PROTEIN. Gene name(s)PSBB. From Hordeum vulgare (Barley), Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

FUNCTION: THIS PROTEIN CONJUGATES WITH CHLOROPHYLL & CATALYZES THE PRIMARY LIGHT-INDUCED PHOTOCHEMICAL PROCESSES OF PHOTOSYSTEM II. SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE. SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

[1] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 89240047. Andreeva A.V., Buryakova A.A., Reverdatto S.V., Chakhmakhcheva O.G., Efimov V.A.; "Nucleotide sequence of the 5.2 kbp barley chloroplast DNA fragment, containing psbB-psbH-petB-petD gene cluster."; Nucleic Acids Res. 17:2859-2860(1989).
[2] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 92207253. Efimov V.A., Andreeva A.V., Reverdatto S.V., Chakhmakhcheva O.G.; "Photosystem II of rye. Nucleotide sequence of the psbB, psbC, psbE, psbF, psbH genes of rye and chloroplast DNA regions adjacent to them."; Bioorg. Khim. 17:1369-1385(1991).
[3] SEQUENCE OF 411-420. Hinz U.G.; "Isolation of the photosystem II reaction center complex from barley. Characterization by cicular dichroism spectroscopy and amino acid sequencing."; Carlsberg Res. Commun. 50:285-298(1985).

1000. QRPTase. Quinolinate phosphoribosyl transferase.
Quinolinate phosphoribosyl transferase (QPRTase) or nicotinate-nucleotide pyrophosphorylase EC:2.4.2.19 is involved in the de novo synthesis of NAD in both

prokaryotes and eukaryotes. It catalyses the reaction of quinolinic acid with 5-phosphoribosyl-1-pyrophosphate (PRPP) in the presence of Mg2+ to give rise to nicotinic acid mononucleotide (NaMN), pyrophosphate and carbon dioxide [1,2]. Number of members: 26.

[1]Medline: 97169443. A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. Eads JC, Ozturk D, Wexler TB, Grubmeyer C, Sacchettini JC; Structure 1997;5:47-58.

[2]Medline: 96139309. The sequencing expression, purification, and steady-state kinetic analysis of quinolinate phosphoribosyl transferase from Escherichia coli. Bhatia R, Calvo KC; Arch Biochem Biophys 1996;325:270-278.

1001. R3H domain

The name of the R3H domain comes from the characteristic spacing of the most conserved arginine and histidine residues. The function of the domain is predicted to be binding ssDNA. Number of members: 28

[1]Medline: 99003905 The R3H motif: a domain that binds single-stranded nucleic acids. Grishin NV; Trends Biochem Sci 1998;23:329-330.

1002. recF protein signatures (RecF)

The prokaryotic protein recF [1,2] is a single-stranded DNA-binding protein which also probably binds ATP. RecF is involved in DNA metabolism; it is required for recombinational DNA repair and for induction of the SOS response. RecF is a protein of about 350 to 370 amino acid residues; there is a conserved ATP-binding site motif 'A' (P-loop) in the N-terminal section of the protein as well as two other conserved regions, one located in the central section, and the other in the C-terminal section. Signature patterns were derived from these two regions.

Consensus pattern [LIVM]-x(4)-[LIF]-x(6)-[LIF]-[LVF]-x-[GE]-[GSTAD]-[PA]- x(2)-R-R-x-[FYW]-[LIVMF]-D Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern[LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L Sequences known to belong to this class detected by the patternALL, except for T. palidum recF.

[1] Sandler S.J., Chackerian B., Li J.T., Clark A.J. Nucleic Acids Res. 20:839-845(1992).
[2] Alonso J.C., Fisher L.M.; Mol. Gen. Genet. 246:680-686(1995).

1003. RibD C-terminal domain (RibD_C)

The function of this domain is not known, but it is thought to be involved in riboflavin biosynthesis. This domain is found in the C terminus of RibD/RibG Swiss:P25539, in combination with dCMP_cyt_deam, as well as in isolation in some archaebacterial proteins Swiss:P95872.
Number of members: 21

1004. Ribosomal protein L16 signatures (Ribosomal_L16)

Ribosomal protein L16 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L16 is known to bind directly the 23S rRNA and to be located at the A site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:
- Eubacterial L16.
- Algal and plant chloroplast L16.
- Cyanelle L16.
- Plant mitochondrial L16.

L16 is a protein of 133 to 185 amino-acid residues. As signature patterns, we selected two conserved regions in the central section of these proteins.

Consensus pattern [KR](2)-x-[GSAC]-[KRQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP] Sequences known to belong to this class detected by the pattern ALL.

Consensus patternR-M-G-x-[GR]-K-G-x(4)-[FWKR] Sequences known to belong to this class detected by the patternALL.

[1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

1005. Ribosomal protein L32e signature (Ribosomal_L32E)

A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L32 [1].
- Drosophila RP49 [2].
- Trichoderma harzianum L32 [3].
- Yeast L32e (YBL092w).
- Archaebacterial L32e [4].

These proteins have 135 to 240 amino-acid residues. As a signature pattern, a stretch of about 20 residues located in the N-terminal part of these proteins was seleced.

Consensus patternF-x-R-x(4)-[KR]-x(2)-[KR]-[LIVMF]-x(3,5)-W-R-[KR]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

[1] Jacks C.M., Powaser C.B., Hackett P.B. Gene 74:565-570(1988).
[2] Aguade M. Mol. Biol. Evol. 5:433-441(1988).
[3] Lora J.M., Garcia I., Benitez T., Llobell A., Pintor-Toro J.A. Nucleic Acids Res. 21:3319-3319(1993).
[4] Arndt E., Scholzen T., Kroemer W., Hatakeyama T., Kimura M. Biochimie 73:657-668(1991).

1006. (Ribosomal_S3) Ribosomal protein S3 signature

PROSITE: PDOC00474. PROSITE cross-reference(s) PS00548; RIBOSOMAL_S3

Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

-Eubacterial S3.
-Algal and plant chloroplast S3.
-Cyanelle S3.
-Archaebacterial S3.
-Plant mitochondrial S3.

-Vertebrate S3.

-Insect S3.

-Caenorhabditis elegans S3 (C23G10.3).

-Yeast S3 (Rp13).

S3 is a protein of 209 to 559 amino-acid residues. A conserved region located in the C-terminal section was selected as a signature pattern.

Consensus pattern[GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS]. Sequences known to belong to this class detected by the patternALL, except for some mitochondrial S3.

[1]Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

1007. RimM - RimM

The RimM protein is essential for efficient processing of 16S rRNA [1]. The RimM protein was shown to have affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14.

[1]Medline: 98083058. RimM and RbfA are essential for efficient processing of 16S rRNA in Escherichia coli. Bylund GO, Wipemo LC, Lundberg LA, Wikstrom PM; J Bacteriol 1998;180:73-82.

1008. RNA_pol_A - RNA polymerase alpha subunit

-!- RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases).

-!- Members of this family include: A subunit from eukaryotes, gamma subunit from cyanobacteria, beta' subunit from eubacteria, A' subunit from archaebacteria, B" from chloroplasts. Number of members: 139.

[1]Medline: 97066998. Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst SA, Goldfarb A; J Biol Chem 1996;271:27969-27974.

1009. RuBisCO_large - Ribulose bisphosphate carboxylase large chain active site
PROSITE: PDOC00142; PROSITE cross-reference(s) PS00157; RUBISCO_LARGE

Ribulose bisphosphate carboxylase (EC 4.1.1.39) (RuBisCO) [1,2] catalyzes the initial step in Calvin's reductive pentose phosphate cycle in plants as well as purple and green bacteria. It consists of a large catalytic unit and a small subunit of undetermined function. In plants, the large subunit is coded by the chloroplastic genome while the small subunit is encoded in the nuclear genome. Molecular activation of RuBisCO by CO2 involves the formation of a carbamate with the epsilon-amino group of a conserved lysine residue. This carbamate is stabilized by a magnesium ion. One of the ligands of the magnesium ion is an aspartic acid residue close to the active site lysine [3]. A pattern was developed which includes both the active site residue and the metal ligand, and which is specific to RuBisCO large chains.

Consensus patternG-x-[DN]-F-x-K-x-D-E [K is the active site residue] [The second D is a magnesium ligand]. Sequences known to belong to this class detected by the patternALL, except for Cheilopleuria biscuspis RuBisCO.

[1]Miziorko H.M., Lorimer G.H. Annu. Rev. Biochem. 52:507-535(1983).
[2]Akazawa T., Takabe T., Kobayashi H. Trends Biochem. Sci. 9:380-383(1984).
[3]Andersson I., Knight S., Schneider G., Lindqvist Y., Lundqvist T., Branden C.-I., Lorimer G.H. Nature 337:229-234(1989).

1010. Rve - Integrase core domain
Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc binding domain Integrase_Zn. This domain is the central catalytic domain. The carboxyl terminal domain that is a non-specific DNA binding domain integrase. The catalytic domain acts as an endonuclease when two nucleotides are removed from the 3' ends of the blunt-ended viral DNA made by reverse transcription. This domain also catalyses the DNA strand transfer reaction of the 3' ends of the viral DNA to the 5' ends of the integration site [1]. Number of members: 694.

[1]Medline: 95099322. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; Science 1994;266:1981-1986.

1011. (SBP_bac_3) Bacterial extracellular solute-binding proteins, family 3 signature PROSITE: PDOC00798. PROSITE cross-reference(s) PS01039; SBP_BACTERIAL_3

Bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. The protein components of these traffic systems include one or two transmembrane protein components, one or two membrane-associated ATP-binding proteins (ABC transporters; see <PDOC00185>) and a high affinity periplasmic solute-binding protein. The later are thought to bind the substrate in the vicinity of the inner membrane, and to transfer it to a complex of inner membrane proteins for concentration into the cytoplasm.

In gram-positive bacteria which are surrounded by a single membrane and have therefore no periplasmic region the equivalent proteins are bound to the membrane via an N-terminal lipid anchor. These homolog proteins do not play an integral role in the transport process per se, but probably serve as receptors to trigger or initiate translocation of the solute throught the membrane by binding to external sites of the integral membrane proteins of the efflux system.

In addition at least some solute-binding proteins function in the initiation of sensory transduction pathways.

On the basis of sequence similarities, the vast majority of these solute-binding proteins can be grouped [1] into eight families of clusters, which generally correlate with the nature of the solute bound.

Family 3 groups together specific amino acids and opine-binding periplasmic proteins and a periplasmic homolog with catalytic activity:

-Histidine-binding protein (gene hisJ) of Escherichia coli and related bacteria. An homologous lipoprotein exists in Neisseria gonorrhoeae.

-Lysine/arginine/ornithine-binding proteins (LAO) (gene argT) of Escherichia coli and related bacteria are involved in the same transport system than hisJ. Both solute-binding proteins interact with a common membrane-bound receptor hisP of the binding protein dependent transport system HisQMP.

-Glutamine-binding proteins (gene glnH) of Escherichia coli and Bacillus stearothermophilus.

-Glutamate-binding protein (gene gluB) of Corynebacterium glutamicum.
-Arginine-binding proteins artI and artJ of Escherichia coli.
-Nopaline-binding protein (gene nocT) from Agrobacterium tumefaciens.
-Octopine-binding protein (gene occT) from Agrobacterium tumefaciens.
-Major cell-binding factor (CBF1) (gene: peb1A) from Campylobacter jejuni.
-Bacteroides nodosus protein aabA.
-Cyclohexadienyl/arogenate dehydratase of Pseudomonas aeruginosa, a periplasmic enzyme which forms an alternative pathway for phenylalanine biosynthesis.
-Escherichia coli protein fliY.
-Vibrio harveyi protein patH.
-Escherichia coli hypothetical protein ydhW.
-Bacillus subtilis hypothetical protein yckB.
-Bacillus subtilis hypothetical protein yckK.

The signature pattern is located near the N-terminus of the mature proteins.
Consensus patternG-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN]
Sequences known to belong to this class detected by the patternALL.

[1]Tam R., Saier M.H. Jr. Microbiol. Rev. 57:320-346(1993).

1012. Sec7 - Sec7 domain
The Sec7 domain is a guanine-nucleotide-exchange-factor (GEF)for the arf family [2].
Number of members: 32.

[1]Medline: 98169075. Structure of the Sec7 domain of the Arf exchange factor. ARNO. Cherfils J, Menetrey J, Mathieu M, Le Bras G, Robineau S, Beraud-Dufour S, Antonny B, Chardin P; Nature 1998;392:101-105.
[2]Medline: 97100951. A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. Chardin P, Paris S, Antonny B, Robineau S, Beraud-Dufour S, Jackson CL, Chabre M. Nature 1996;384:481-484.

1013. SecA_protein. SecA protein, amino terminal region

SecA protein binds to the plasma membrane where it interacts with proOmpA to support translocation of proOmpA through the membrane. SecA protein achieves this translocation, in association with SecY protein, in an ATP dependent manner. SecA possesses the ATPase activity. The carboxyl terminus has similarity with the helicase carboxyl terminus. See Ribosomal_L5. Number of members: 45.

[1]Medline: 98309858. Amino-terminal region of SecA is involved in the function of SecG for protein translocation into Escherichia coli membrane vesicles. Mori H, Sugiyama H, Yamanaka M, Sato K, Tagaya M, Mizushima S; J Biochem (Tokyo) 1998;124:122-129.
[2]Medline: 89251629. SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. Lill R, Cunningham K, Brundage LA, Ito K, Oliver D, Wickner W; EMBO J 1989;8:961-966.

1014. Seedstore_2S - 2S seed storage family
Members of this family are composed of two chains (both included in the alignment), these are co-translated and later cleaved. The two chains are disulphide linked together. Number of members: 27.

[1]Medline: 97121264. 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. Rico M, Bruix M, Gonzalez C, Monsalve RI, Rodriguez R; Biochemistry 1996;35:15672-15682.

1015. Smr - Smr domain
This family includes the Smr (Small MutS Related) proteins, and the C-terminal region of the MutS2 protein. It has been suggested that this domain interacts with the MutS1 Swiss:P23909 protein in the case of Smr proteins and with the N-terminal MutS related region of MutS2 Swiss:P94545 [1]. Number of members: 14.

[1]Medline: 10431172. Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. Moreira D, Philippe H; Trends Biochem Sci 1999;24:298-300.

1016. (SSF) Sodium:solute symporter family signatures and profile
PROSITE: PDOC00429. PROSITE cross-reference(s)PS00456; NA_SOLUT_SYMP_1 PS00457; NA_SOLUT_SYMP_2 PS50283; NA_SOLUTE_SYMP_3

It has been shown [1,2] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:solute symporter family (SSF) and currently consists of the following proteins:

-Mammalian Na+/glucose co-transporter.

-Mammalian Na+/myo-inositol co-transporter.

-Mammalian Na+/nucleoside co-transporter.

-Mammalian Na+/neutral amino acid co-transporter.

-Escherichia coli Na+/proline symporter (gene putP).

-Escherichia coli Na+/pantothenate symporter (gene panF).

-Escherichia coli hypothetical protein yidK.

-Escherichia coli hypothetical protein yjcG.

-Bacillus subtilis hypothetical protein ywcA (ipa-31R).

These integral membrane proteins are predicted to comprise at least ten membrane spanning domains. Two conserved regions were selected as signature patterns; the first one is located in the fourth transmembrane region and the second one in a loop between two transmembrane regions in the C-terminal part of these proteins.

Consensus pattern[GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-[SAP]. Sequences known to belong to this class detected by the patternALL.
Consensus pattern[GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-x-[LIVMGA] Sequences known to belong to this class detected by the patternALL, except for E.coli yidK.
Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

[1]Reizer J., Reizer A., Saier M.H. Jr. Res. Microbiol. 141:1069-1072(1991).
[2]Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).

1017. SurE - Survival protein SurE

E. coli cells with the surE gene disrupted are found to survive poorly in stationary phase [1]. It is suggested that SurE may be involved in stress response. Yeast also contains a member of

the family Swiss:P38254. Swiss:P30887 can complement a mutation in acid phosphatase, suggesting that members of this family could be phosphatases. Number of members: 17.

[1]Medline: 95014035. A new gene involved in stationary-phase survival located at 59 minutes on the Escherichia coli chromosome. Li C, Ichikawa JK, Ravetto JJ, Kuo HC, Fu JC, Clarke S; J Bacteriol 1994;176:6015-6022.
[2]Medline: 93046805. Complementation of Saccharomyces cerevisiae acid phosphatase mutation by a genomic sequence from the yeast Yarrowia lipolytica identifies a new phosphatase. Treton BY, Le Dall MT, Gaillardin CM; Curr Genet 1992;22:345-355.

1018. Synuclein - Synuclein

There are three types of synucleins in humans, these are called alpha, beta and gamma. Alpha synuclein has been found mutated in families with autosomal dominant Parkinson's disease. A peptide of alpha synuclein has also been found in amyloid plaques in Alzheimer's patients. Number of members: 12.

[1]Medline: 98424410. The synuclein family. Lavedan C; Genome Res 1998;8:871-880.

1019. (T-box) T-box domain signatures

PROSITE: PDOC00972. PROSITE cross-reference(s) PS01283; TBOX_1 PS01264; TBOX_2

A number of eukaryotic DNA-binding proteins contain a domain of about 170 to 190 amino acids known as the T-box domain [1,2,3] and which probably binds DNA. The T-box has first been found in the mice T locus (Brachyury) protein, a transcription factor involved in mesoderm differentiation. It has since been found in the following proteins:

-Vertebrate and invertebrate homologs of the T protein.

-Mammalian proteins TBX1 to TBX6.

-Mammalian protein TBR1 which is expressed specifically in brain.

-Xenopus laevis eomesodermin (eomes).

-Xenopus laevis Vegt (or Antipodean), a transcription factor that activates the expression of wnt-8, eomes and Brachyury.

-Chicken TbxT.

-Drosophila protein optomotor-blind (omb).

-Drosophila protein brachyenteron (byn) (also known as Trg), which is

required for the specification of the hindgut and anal pads.

-Drosophila protein H15.

-Caenorhabditis elegans protein tbx-12.

-Caenorhabditis elegans hypothetical proteins F21H11.3, F40H6.4, T07C4.2, T07C4.6 and ZK177.10.

Two conserved regions were selected as signature patterns for the T-domain. The first region corresponds to the N-terminal of the domain and the second one to the central part.

Consensus patternL-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ]

Sequences known to belong to this class detected by the patternALL, except for C.elegans ZK177.10.

Consensus pattern[LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x- F

Sequences known to belong to this class detected by the patternALL, except for C.elegans tbx-12, ZK177.10 and Drosophila H15.

[1]Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M., Silver L.M. Nat. Genet. 7:383-389(1994).

[2] Agulnik S.I., Garvey N., Hancock S., Ruvinsky I., Chapman D.L., Agulnik I., Bollag R.J., Papaioannou V.E., Silver L.M. Genetics 144:249-254(1996).

[3]Papaioannou V.E. Trends Genet. 13:212-213(1997).

1020. Toprim - Toprim domain

This is a conserved region from DNA primase. This corresponds to the Toprim domain common to DnaG primases, topoisomerases, OLD family nucleases and RecR proteins [1]. Both DnaG motifs IV and V are present in the alignment, the DxD (V) motif may be involved in Mg2+ binding and mutations to the conserved glutamate (IV) completely abolish DnaG type primase activity [1]. DNA primase EC:2.7.7.6 is a nucleotidyltransferase it synthesizes the oligoribonucleotide primers required for DNA replication on the lagging strand of the replication fork; it can also prime the leading stand and has been implicated in cell division [2]. Number of members: 133.

[1]Medline: 98391745. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Aravind L, Leipe DD, Koonin EV; Nucleic Acids Res 1998;26:4205-4213.

[2]Medline: 97368180. Cloning and analysis of the dnaG gene encoding Pseudomonas putida DNA primase. Szafranski P, Smith CL, Cantor CR; Biochim Biophys Acta 1997;1352:243-248.

[3]Medline: 94124015. The Haemophilus influenzae dnaG sequence and conserved bacterial primase motifs. Versalovic J, Lupski JR; Gene 1993;136:281-286.

1021. TraB - TraB family

pAD1 is a hemolysin/bacteriocin plasmid originally identified in Enterococcus faecalis DS16. It encodes a mating response to a peptide sex pheromone, cAD1, secreted by recipient bacteria. Once the plasmid pAD1 is acquired, production of the pheromone ceases--a trait related in part to a determinant designated traB. However a related protein is found in C. elegans Swiss:Q94217, suggesting that members of the TraB family have some more general function. Number of members: 12.

[1]Medline: 94302142. Characterization of the determinant (traB) encoding sex pheromone shutdown by the hemolysin/bacteriocin plasmid pAD1 in Enterococcus faecalis. An FY, Clewell DB; Plasmid 1994;31:215-221.

1022. (Transpo_mutator) Transposases, Mutator family, signature

PROSITE: PDOC00770. PROSITE cross-reference(s) PS01007; TRANSPOSASE_MUTATOR

Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2,3,E1] to consist of transposases from the following elements:

-Mutator from Maize.

-Is1201 from Lactobacillus helveticus.

-Is905 from Lactococcus lactis.

-Is1081 from Mycobacterium bovis.

-Is6120 from Mycobacterium smegmatis.

-Is406 from Pseudomonas cepacia.

-IsRm3 from Rhizobium meliloti.

-IsRm5 from Rhizobium meliloti.

-Is256 from Staphylococcus aureus.

-IsT2 from Thiobacillus ferrooxidans.

The maize Mutator transposase (MudrA) is a protein of 823 amino acids; the bacterial transposases listed above are proteins of 300 to 420 amino acids. These proteins contain a conserved domain of about 130 residues; a signature pattern was derived from the most conserved part of this domain.

Consensus patternD-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-H. Sequences known to belong to this class detected by the patternALL.

[1]Eisen J.A., Benito M.-I., Walbot V. Nucleic Acids Res. 22:2634-2636(1994).

[2]Guilhot C., Gicquel B., Davies J., Martin C. Mol. Microbiol. 6:107-113(1992).

[3]Wood M.S., Byrne A., Lessie T.G. Gene 105:101-105(1991).

1023. Transposase_8 - Transposase

Transposase proteins are necessary for efficient DNA transposition. This family consists of various E. coli insertion elements and other bacterial transposases some of which are members of the IS3 family. Number of members: 58.

[1]Medline: 97324595. Genetic organization and transposition properties of IS511. D. A. Mullin, D. L. Zies, A. H. Mullin, N. Caballera & B. Ely; Mol Gen Genet 1997;254:456-463.

[2]Medline: 97128810. The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of Streptomyces lividans 66. J. Fischer, H. Maier, P. Viell & J. Altenbuchner; Gene 1996;180:81-89.

[3]Medline: 97074647. Identification and nucleotide sequence of Rhizobium meliloti insertion sequence ISRm6, a small transposable element that belongs to the IS3 family. S. Zekri & N. Toro; Gene 1996;175:43-48.

1024. tRNA_int_endo - tRNA intron endonuclease

Members of this family cleave pre tRNA at the 5' and 3' splice sites to release the intron EC:3.1.27.9. Number of members: 8.

[1]Medline: 97344075. Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Kleman-Leyer K, Armbruster DW, Daniels CJ; Cell 1997;89:839-847.

1025. Urease - Urease signatures

PROSITE: PDOC00133PROSITE cross-reference(s) PS01120; UREASE_1 PS00145; UREASE_2

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine was selected.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel].Sequences known to belong to this class detected by the patternALL.
Consensus pattern[LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]. Sequences known to belong to this class detected by the patternALL.

[1]Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).
[2]Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).
[3]Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

1026. Urease_beta - Urease beta subunit.
This subunit is known as alpha in Heliobacter. Number of members: 35.

[1]Medline: 95273988. The crystal structure of urease from Klebsiella aerogenes. Jabri E, Carr MB, Hausinger RP, Karplus PA; Science 1995;268:998-1004.

1027. UvrD-helicase - UvrD/REP helicase

The Rep family helicases are composed of four structural domains. The Rep family function as dimers. REP helicases catalyse ATP dependent unwinding of double stranded DNA to single stranded DNA. Swiss:P23478, Swiss:P08394 have large insertions near to the carboxy-terminus relative to other members of the family. Number of members: 52.

[1] Medline: 97433075. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G; Cell 1997;90:635-647.

1028. V-type ATPase 116kDa subunit family (V_ATPase_sub_a)

This family consists of the 116kDa V-type ATPase (vacuolar (H+)-ATPases) subunits, as well as V-type ATP synthase subunit i. The V-type ATPases family are proton pumps that acidify intracellular compartments in eukaryotic cells for example yeast central vacuoles, clathrin-coated and synaptic vesicles. They have important roles in membrane trafficking processes [1]. The 116kDa subunit (subunit a) in the V-type ATPase is part of the V0 functional domain responsible for proton transport. The a subunit is a transmembrane glycoprotein with multiple putative transmembrane helices t has a hydrophilic amino terminal and a hydrophobic carboxy terminal [1,2]. It has roles in proton transport and assembly of the V-type ATPase complex [1,2]. This subunit is encoded by two homologous gene in yeast VPH1 and STV1 [2].
Number of members: 27

[1] Forgac M; Medline: 99240666 Structure and properties of the vacuolar (H+)-ATPases." J Biol Chem 1999;274:12951-12954.
[2] Forgac M; Medline: 99270697 Structure and properties of the clathrin-coated vesicle and yeast vacuolar V-ATPases." J Bioenerg Biomembr 1999;31:57-65.

1029. Viral (Superfamily 1) RNA helicase (Viral_helicase1)
Number of members: 260

[1] Koonin EV, Dolja VV; Medline: 94094568 Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." Crit Rev Biochem Mol Biol 1993;28:375-430.

1030. Vesicular monoamine transporter (VMAT)

This family consists of various vesicular amine transporters with 12 transmembrane helices. These included vesicular acetylcholine transporters (VAChT) [3], and vesicular monoamine transporters (VMATs) [1,2] isoforms 1 adrenal and 2 brain (VMAT1 and VMAT2).

These proteins transport biogenic amines into synaptic vesicles or chromaffin granules [4]. VMATs pack monoamine neurotransmitters into secretary vesicles for regulated exocytotic release, they also protect against the parkinsonian neurotoxins MPP+ by transporting it into vesicles preventing it from acting on mitochondria [1].

Also in the family is C. elegans UNC-17 a putative vesicular acetylcholine transporter mutations in UNC-17 cause impaired neuromuscular function, giving rise to jerky or uncoordinated movement, [4].
Number of members: 15

[1] Krantz DE, Peter D, Liu Y, Edwards RH; Medline: 97197857 Phosphorylation of a vesicular monoamine transporter by casein kinase II." J Biol Chem 1997;272:6752-6759.
[2] Erickson JD, Varoqui H, Schafer MK, Modi W, Diebler MF, Weihe E, Rand J, Eiden LE, Bonner TI, Usdin TB; Medline: 94350930 Functional identification of a vesicular acetylcholine transporter and its expression from a 'cholinergic' gene locus." J Biol Chem 1994;269:21929-21932.
[3] Erickson JD, Schafer MK, Bonner TI, Eiden LE, Weihe E; Medline: 96209876 Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter." Proc Natl Acad Sci U S A 1996;93:5166-5171.
[4] Alfonso A, Grundahl K, Duerr JS, Han HP, Rand JB; Medline: 3342494 The Caenorhabditis elegans unc-17 gene: a putative vesicular acetylcholine transporter." Science 1993;261:617-619.

1031. WW/rsp5/WWP domain signature and profile. Cross-reference(s): PS01159; WW_DOMAIN_1; PS50020; WW_DOMAIN_2

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

--Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.

--Utrophin, a dystrophin-like protein of unknown function.

--Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].

--Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].

--Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>), followed by a histidine-rich region, 3 WW domains and a HECT domain.

--Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

--Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).

--Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.

--IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

--Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2-type myosin, each containing two WW-domains at the N-terminus.

--Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.

--Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Description of pattern(s) and/or profile(s):

Consensus patternW-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

[1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).

[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).

[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).

[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).

[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).

[6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).

1032. XPA protein signatures. cross-reference(s): XPA_1 PROSITE PS00752; PS00753;XPA_2.

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due

to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. XP-A is the most severe form of the disease and is due to defects in a 30 Kd nuclear protein called XPA (or XPAC) [2].

The sequence of the XPA protein is conserved from higher eukaryotes [3] to yeast (gene RAD14) [4]. XPA is a hydrophilic protein of 247 to 296 amino-acid residues which has a C4-type zinc finger motif in its central section.

Two signature were developed patterns for XPA proteins. The first corresponds to the zinc finger region, the second to a highly conserved region located some 12 residues after the zinc finger region.

Consensus patternC-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C
Consensus pattern[LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE]

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).
[2] Miura N., Miyamoto I., Asahina H., Satokata I., Tanaka K., Okada Y. J. Biol. Chem. 266:19786-19789(1991).
[3] Shimamoto T., Kohno K., Tanaka K., Okada Y. Biochem. Biophys. Res. Commun. 181:1231-1237(1991).
[4] Bankmann M., Prakash L., Prakash S. Nature 355:555-558(1992).

1033. YCF9

This family consists of the hypothetical protein product of the YCF9 gene from chloroplasts and cyanobacteria. Number of members: 16

1034. (DUF15)

It is highly conserved between eubacteria and eukaryotes.

Number of members: 30

1035. Lumenal portion of Cytochrome b559, alpha (gene psbE) subunit. (cytochr_b559a)

This family is the lumenal portion of cytochrome b559 alpha chain, matches to this family should be accompanied by a match to the cytochr_b559 family also. The Prosite pattern pattern matches the transmembrane region of the cytochrome b559 alpha and beta subunits.
Number of members: 16

A. Asparaginase 2

Asparaginase II (L-asparagine aminohydrolase II) is an extracellular protein that may be associated with the cell wall and whose expression is affected by the availability of nitrogen. Asparaginase II catalyzes the reaction of L-Asparagine + H_2O = L-Aspartate + NH_3. As many leukemias have high requirements for aspartic acid, asparaginase II proteins are useful as reagents for screening compounds for activity as leukemia chemotherapy products. Asparaginase II protein can also be over- or under-expressed to alter amino acid content in plant tissues or to modify nitrogen fixation and/or nitrogen metabolism in plants.

Ref: Bon et al. (1997) Appl Biochem Biotechnol 63-65: 203-12

B. Chloroa b-bind

Chlorophyll a-b binding proteins are located in the thylakoid membranes of the chloroplast and bind chlorophyll a and chlorophyll b, thereby triggering a chemical reaction (photosynthesis). These proteins are useful in controlling the rate, efficiency and/or output of photosynthesis. Overexpression of chlorophyll a-b binding proteins is expected to increase the rate of photosynthesis.

Ref: Leutwiler et al. (1986) Nucleic Acids Res 14: 4051-64
Brandt et al. (1992) Plant Mol Biol 19: 699-703

C. DMRL synthase

DMRL Synthase (6,7-Dimethyl-8-Ribityllumazine Synthase) catalyzes the last step in riboflavin (Vitamin B_2) synthesis, condensing 5-amino-6-(1'-D)-ribityl-amino-2,4(1H, 3H)-Pyrimidinedione with L-3,4-Dihydroxy-2-Butanone 4-Phosphate producing 6,7-Dimethyl-8-(1-D-Ribityl)Luminazine . The enzyme forms a homopentamer. Engineering of these proteins or those with homologous sequences/structures may allow control of the amounts of vitamin B_2 available in plants and/or accumulation of pigment, as well as altering reactions requiring hydrogen ion carriers/transmitters.

Ref: Garcia-Ramirez et al. (1995) J Biol Chem **270**: 23801-7

D. E1_N

These proteins are ATP-dependent DNA helicases that are required for initiation of viral DNA replication. They form a complex with the viral E2 protein. The E1-E2 complex binds to the replication origin that contains binding sites for both proteins. The majority of sequences known for this group of proteins are from various papillomaviruses, a type of double stranded DNA virus. In plants, the prototype double stranded DNA virus is Cauliflower Mosaic virus (CaMV). Manipulation of these proteins, especially to produce variant proteins that form non-productive complexes, enables production of plants that are resistant to infection by double stranded DNA viruses.

Ref: Yang et al. (1993) PNAS USA **90**: 5086-90
Ustav and Stenlund (1991) EMBO J **10**: 449-57
Callaway et al. (1996) Mol Plant Microbe Interact 9: 810-8

E. EF1_G

Elongation Factor-1 is composed of four subunits: alpha, beta, delta and gamma. Gamma subunits are presumed to play a role in anchoring the complex to other cellular components. Studies of EF-1 genes in plants suggests that different forms of the EF-1 subunits may be expressed in particular organs or in response to stress. Manipulation of the activity of these proteins, either by altered expression level or by structural mutation, may result in the accumulation of a particular protein in a chosen organ or allow production of particular proteins during stress conditions.

Ref: Kinzy et al. (1994) NAR 22: 2703-7
Dunn et al. (1993) Plant Mol Biol 23: 221-5
Aguilar et al. (1991) Plant Mol Biol 17: 351-60

F. ENV_polyprotein

This family comprises the envelope or coat proteins known from a number of different retroviruses. In mammalian species, retroviruses are responsible for diseases such as leukemia and HIV. In plants, retroviruses are known in both monocot (e.g. Zeon-1) and dicot (e.g. Arabidopsis and tobacco) species and have been shown to induce mutant alleles at new loci. Engineering of plant ENV proteins may allow mobilization or targeting of endogenous or introduced retroviruses, in essence generating a new method for mutant production, gene tagging and the like.

Ref: Mamoun et al (1990) J Virol 64: 4180-8
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

G. Glycosyl_hydr9

Proteins having this domain (previously known as the glycosyl hydrolase family 5 domain) catalyze the endohydrolysis of 1,4-β-D-glucosidic linkages in cellulose. Numerous plant proteins with this domain exist and are expressed in an organ specific manner. They are involved in the fruit ripening process, in cell elongation and plant reproduction. Modulation of the activity of these proteins, either by over- or under-expression or by mutation of the polypeptide, could be used to affect post-harvest physiology (e.g. rate of ripening) or for engineering reproductive sterility.

Ref: Giorda et al. (1990) Biochemistry 29: 7264-9
Tucker et al. (1988) Plant Physiol 88: 1257-62
Shani et al. (1997) 43: 837-42

Milligan and Gasser (1995) Plant Mol Biol 28: 691-711

H. Glycosyl_hydr14

The β-amylases (family 14 of glycosyl hydrolases) catalyze the hydrolysis of 1,4-α-glucosidic linkages in polysaccharides and remove successive maltose units from the non-reducing ends of the chains. Mutants of β-amylase in Arabidopsis exhibited altered degradation of starch throughout the diurnal cycle. In addition, the mutant phenotypes indicated that these enzymes not only affect carbohydrate metabolism/catabolism, but also influence the amount of pigment stored within particular cells. Manipulation of the β-amylase genes enables control of plant pigmentation (for example, fibre pigment in cotton) as well as carbohydrate synthesis and degradation.

Ref: Zeeman et al. (1998) Plant J 15: 357-65
Hirano and Nakamura (1997) Plant Physiol 114: 5675-82
Kitamoto et al. (1988) J Bacteriol 170: 5848-54

I. Glycosyl_hydr15

Glycosyl hydrolases from family 15 (such as 1,4-Alpha-D-Glucan glucohydrolase,) catalyze the hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from the non-reducing ends of the chains resulting in the release of β-D-Glucose. In plants these proteins have been tied to the mobilization of the xyloglucan stored in the cotyledonary cell walls. Proteins such as these could be varied to affect the rate of plant growth (for example during germination), storage and/or use of glucose and other sugars by plant tissues and alteration of the properties, such as elasticity, of plant cell walls.

Ref: Crombie et al. (1998) Plant J 15: 27-38
Hata et al. (1991) Agric Biol Chem 55: 941-9

J. Glycosyl_hydr20

Members of the family 20 glycosyl hydrolases catalyze the hydrolysis of terminal non-reducing N-acetly-D-hexosamine residues in N-acetyl-β-D-hexosaminides. N-acetyl-β -glucosaminidase belongs to this family and exists in several different forms (consisting of various combinations of alpha and beta chains) depending on the organism. Family 20 glycosyl hydrolases have been implicated in lysosomal storage diseases (such as Sandhoff disease) and glycogen storage disease in humans. These types of proteins are also responsible for the hydrolysis of chitin. In plants, these proteins could be useful in controlling carbohydrate catabolism, thereby influencing the amount of sugars available for storage and/or use in other metabolic pathways. In addition, it is possible that such proteins could be used to engineer an endogenous insect protection mechanism, e.g. by secretion of a chitin-hydrolyzing composition by the plant.

Ref: Graham et al (1988) J Biol Chem 263: 16823-9
O'Dowd et al. (1988) Biochemistry 27: 5216-26

K. HMG box

The HMG box is a novel type of DNA-binding domain found in a diverse group of proteins. Numerous plant proteins contain this domain, such as the HMG1/2-like proteins. The expression of some of these HMG proteins appears to be regulated by circadian rhythms and in a light dependent manner, occurring at higher levels in roots, for example and lower levels in light-grown tissues such as cotyledons. Generally, HMG proteins are thought to influence transcription regulation. In plants, HMGs are believed to have a role in maintaining patterns of circadian-regulated expression for other genes, suggesting that these proteins could be exploited to control growth and development.

Ref: Laudet et al. (1993) Nucleic Acids Res 21: 2493-501
Zheng et al. (1993) Plant Mol Biol 23: 813-23
Grasser et al. (1993) Plant Mol Biol 23: 619-25

L. IL2

Interleukin-2 (IL-2)is produced in mammals by T cells in response to antigenic or mitogenic stimulation and is crucial for proper regulation and functioning of the immune response. IL-2 is capable of stimulating B cells, monocytes, lymphokine-activated killer cells, natural killer cells and glioma cells. Plant extracts have also been shown to stimulate the immune system (for example, mistletoe therapy for human cancer). It is known that IL-2 is involved in feedback inhibition pathways that impact the inflammatory response as well as the growth inhibition of tumor reactive T cells. Plant proteins containing IL-2-like sequences are useful as immunity-based therapeutics, acting in a manner similar to IL-2 in mammals.

Ref: Heike et al. (1997) Scand J Immunol 45: 221-6
Ariel et al. (1998) J Immunol 161: 2465-72
Schink (1997) Anticancer Drugs 8 Suppl 1: S47-51

M. Oxidored_FMN

NADPH dehydrogenases catalyze the reaction NADPH + acceptor = NADP(+) + reduced acceptor. One member of this family is yeast old yellow enzyme" (OYE) and is thought to be involved in oxylipin metabolism. A second yeast family member is a protein that binds estrogen binding protein (EBP) in addition to exhibiting oxidoreductase activity. An Arabidopsis homolog to OYE has been described and estrogen binding proteins in plants have been reported. Plant proteins from this class have the potential to be used to modify lipid metabolism/catabolism. These proteins may also have use as therapeutics for breast and prostate cancer, and other abnormal growth in steroid-sensitive tissues.

Ref: Baker et al. (1998) Proc Soc Exp Biol Med 217: 317-21
Schaller and Weiler (1997) J Biol Chem 272: 28066-72
Mandani et al. (1994) PNAS USA 91: 922-6

N. Oxidored_q2

The NADH-plastoquinone oxidoreductases catalyze the reaction NADH + plastoquinone = NAD(+) + plastoquinol. In plants these reactions occur in the chloroplast and are believed to participate in a chloroplast respiratory system. Here, the NDH complex is postulated to act as

a valve to remove excess reduction equivalents in the chloroplasts. Manipulation of these proteins may improve the rate or efficiency of photosynthesis.

Ref: Burrows et al. (1998) EMBO J 17: 868-76
Kofer et al (1998) Mol Gen Genet 258: 166-73
Maier et al. (1995) J Mol Biol 251: 614-28

O. PABP

Polyadenylate binding proteins bind the poly (A) tail of mRNA. Plants, as exemplified by Arabidopsis, contain numerous PABP genes that are expressed in an organ-specific manner. For example, PABP2 is functional in roots and shoots, while PABP5 is expressed predominantly in immature flowers. The PABP proteins are implicated in numerous aspects of posttranscriptional regulation including mRNA turnover and translational initiation. Control of activity of PABP proteins provides the ability to control the expression of various genes in particular organs during development.

Ref: Hilson et al (1993) Plant Physiol 103: 525-33
Belostotsky and Meagher (1993) PNAS USA 90: 6686-90

P. Parvo coat

Parvoviruses are linear single-stranded DNA viruses that are encapsulated by three capsid proteins. Plants are susceptible to infection by single stranded DNA viruses such as Maize streak virus (MSV) and various Gemini viruses. The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of MSV is thought to be involved in intra- and inter-cellular movement within the plant. Engineering of proteins having similarity to parvoviral coat proteins, especially to produce proteins that interfere with maturation of the virus particle, enables the production of plants having better resistance to natural plant single-stranded DNA viruses.

Ref: Liu et al. (1997) J Gen Virol 78: 1265-70
Rohde et al. (1990) Virology 176: 648-51

Q. Pkinase_C

Plant serine/threonine protein kinases possessing this domain are expressed in all tissues and are known to undergo serine-specific autophosphorylation and specifically phosphorylate two ribosomal proteins, P14 and P16. During development, these proteins predominate during high metabolic activity in growing buds, root tips, leaf margins and germinating seeds. They are thought to be involved in the control of plant growth and development. In addition, two genes encoding proteins from this family have been described that help plant cells adapt during cold or high salt stresses. Consequently, engineering Pkinase C proteins provides a way to control general growth/development of the plant as well as a means to provide endogenous protection against environmental stresses.

Ref: Zhang et al. (1994) J Biol Chem 269: 17586-92
Mizoguchi et al. (1995) FEBS Lett 358: 199-204

R. REV

The REV proteins act post-transcriptionally to relieve negative repression of GAG and ENV production in retroviruses such as Human Immounodeficiency Virus type I (HIV-1). Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutations at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant REV proteins enables control of transposition frequencies of corresponding transposable elements and provides a new tool for genetic engineering of plants.

Ref: Sodroski et al. (1986) Nature 321: 412-7
Franchini et al. (1989) PNAS USA 86: 2433-7
Marquet et al. (1995) 77: 113-24
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

S. RuBisCo small

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCo) catalyzes the initial step in the C3 photosynthetic carbon reduction cycle, adding carbon dioxide to D-ribulose 1,5-bisphosphate to form two molecules of 3-phospho-D-glycerate. RuBisCo is comprised of two subunits, one large which is synthesized in the chloroplast, and one small which is synthesized in the cytoplasm and then transported in to the chloroplast. The expression of the small subunit of RuBisCo is light regulated. Manipulation of these proteins could increase the efficiency of photosynthesis or allow alterations in developmental timing.

Ref: Giuliano et al. (1988) PNAS USA 85: 7089-93
Dedonder et al. (1993) Plant Physiol 101: 801-8

T. Sialyltransf

Members of the CMP-N-acetylneuraminate-β-galactosamide-α-2,3-sialyltransferase family catalyze the following reaction:
CMP-N-acetylneuraminate + β-D-galactosyl-1,3-N-acetyl-α-D-galactosaminyl-R = CMP + α-N-acetylneraminyl-2,3-β-D-galactosyl-1,3-N-acetyl-alpha-D-galactosaminyl-R. These proteins are though to be responsible for the synthesis of the sequence neurac-α-2,3-gal-β-1,3-galnac- found on sugar chains)-linked to threonine or serine and also as a terminal sequence on certain gangliosides in mammalian cells. In plants, glycosyltransferases in the Golgi apparatus synthesize cell wall polysaccharides and elaborate the complex glycans of glycoproteins. Engineering of plant sialyltransferases allows targeting of proteins to particular cellular locations or enables the making of changes in cell wall structure.

Ref: Wee et al. (1998) Plant Cell 10: 1759-68
Lee et al. (1994) J Biol Chem 269: 10028-33
Kitagawa and Paulson (1994) J Biol Chem 269: 1394-401

U. Signal

Many plant proteins in this family contain sequences similar to those found in both components of the prokaryotic family of signal transducers known as the two-component systems. This suggests that activation may require a transfer of a phosphate group between

the transmitter domain and the receiver domain. One family member in Arabidopsis appears to be involved in ethylene (a plant hormone) signal transduction. Other proteins in this family appear to be involved in the regulation of gene transcription under conditions of environmental stress. Signal proteins can be exploited to affect plant growth and development and/or control plant responses to stress conditions such as cold, nutrient availability, etc.

Ref: Chang et al. (1993) Science 262: 539-44
Nagaya et al. (1993) Gene 131: 119-124
Gottfert et al. (1990) PNAS USA 87: 2680-4

V. vMSA

vMSA proteins are major surface antigens presenting on the envelope of various retroviruses. Surface antigens of retroviruses are often involved in tropism of the virus. Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutants at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant vMSA proteins enables control of tropism of plant retroviruses that might be used for genetic engineering tools, thus enabling targeting of the virus to particular species and/or tissues of plants.

Ref: Okamoto et al. (1988) J Gen Virol 69: 2575-83
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

W. zf-CCCH

This family of proteins is defined by having two CX(8)CX(5)CX(3)H-type zinc finger domains. These proteins cover a broad range of functions. For example, the COP1 protein acts as a repressor of photomorphogenesis in darkness; light stimuli abolish this suppressive action. In addition, COP1 protein can function as a negative transcriptional regulator capable of direct interaction with components of the G-protein signaling pathway. As a second example, a zf-CCCH protein identified in Arabidopsis appears to be involved in the resistance to DNA damage induced by UV light and chemical DNA-damaging agents.

Overexpression of this class of proteins permits production of plants that are better suited to adverse environments. Manipulation of expression of zf-CCCH proteins functioning as transcriptional regulators, such as COP1, enables manipulation of some signal transduction pathways.

Ref: Pang et al. (1993) Nucleic Acids Res 21: 1647-53
Deng et al. (1992) Cell 71: 791-801

X. zf-RanBP

Proteins falling within this category contain many X-X-F-G and X-F-X-F-G repeats, and may contain RANBP1-like or PPIase domains. Plant proteins having domains similar to these include PAS1 and GMSTI. PAS1 has been shown to have dramatic developmental affects that appear to be correlated with both cell division and cell wall elongation. GMSTI has high identity to the yeast STI stress-inducible gene and has been shown to be heat inducible. Proteins such as these may be useful for controlling growth and form of development.

Ref: Vittorioso et al. (1998) Mol Cell Biol 18: 3034-43
Hernandez Torres et al. (1995) 27: 1221-6

Y. Peptidase M48.

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are located in the membranes of the endoplasmic reticulum. They function in NH_2-terminal proteolytic processing, as shown for the yeast STE24 gene product. This gene is required for the correct processing of α-factor, a yeast pheromone. Family M48 peptidases also appear to be required for some prenylation reactions, mediating COOH-terminal CAAX processing. Prenylation reactions are believed to be involved in the regulation of protein-protein and protein-membrane interactions. As an example, RAS GTPase activity is regulated in part by localization to the inner side of the plasma membrane upon prenylation. In plants, proteins from this family could be involved in pollen-stigma interactions such as those mediating self-pollenation vs. outcrossing, or could be members of several secondary metabolism pathways.

Ref: Fujimura-Kamada et al. (1997) J Cell Biol. 136: 271-85. Tam et al. (1998) J Cell Biol. 142: 635-49.

Z. DNA Pol Viral N

The DNA pol Viral N domain is located at the N-terminal region of DNA polymerase isolated from several retroid viruses such as the Cauliflower Mosaic Virus. The domain motif has also been found in numerous other species from humans to cyanobacteria. In these organisms, this motif seems to be associated with two types of sequences; retrotransposons and mitochondrial genes. In the mitochondrial sequences this domain is potentially involved in the self-splicing conducted by group II introns. Various manipulations of this gene in plants allows control of the numerous retrotransposons endogenous to plant genomes or allows engineering of mitochondrial function, especially to increase efficiency of energy utilization by cells.

REF: Chapdelaine and Bonen (1991) Cell 65: 465-72
Ferat and Miche (1993) Nature 364: 358-61
Wilson et al. (1994) 368: 32-8
Cambareri et al. (1994) 242: 658-65
Gaardner et al. (1981) NAR 9: 2871-2888
Cummings et al. (1990) Curr Genet 17: 375-402
Hattori et al. (1986) Nature 321: 625-8

Aa. Calpain_inhib

This domain is found in calpastatin, an inhibitor protein specific for calpain. Calpain is a non-lysosomal calcium-dependent intracellular protease that appears to be involved in the dynamic changes of the cytoskeleton, especially actin-related structures, during early *Drosophila* embryogenesis [1]. Calpastatins co-exist in cells with calpains and the subcellular distribution of calpastatin is thought to be important to calpain regulation [2]. In plants calpains and calpastatins could be involved in embryogenesis and non-embryogenic organ reiteration. Mutations occurring in calpain inhibitor repeat domains would produce developmental abnormalities such as abnormal leaf, root or flower development.

Refs

1 Emori Y and Saigo K (1994) J Biol Chem 269: 25137-42.

2 Mellgren RL, Lane RD, Mericle MT (1989) Biochim Biophys Acta 999: 71-77.

Ab. chorismate_bind

Chorismate binding domains are present in plant anthranilate synthase (AS) genes. AS genes catalyze the first step in the biosynthesis of tryptophan by converting chorismate and L-glutamine to anthranilate, pyruvate and L-glutamate. Some of these genes are involved in feedback inhibition by tryptophan [1] while some are feedback insensitive [2]. In Arabidopsis, two AS genes have overlapping, but different distributions. One of these AS genes is induced by wounding and bacterial pathogen infiltration [1]. Mutations in the chorismate binding domain would affect the production of tryptophan and could influence the plant's defense system. AS gene products can be used for *in vitro* synthesis of tryptophan and tryptophan derivatives.

Refs

1 Niyogi KK, Fink GR (1992) Plant Cell 4: 721-33.

2 Song HS, Brotherton JE, Gonzales RA, Wilholm JM (1998) Plant Physiol 117:533-43.

Ac. late_protein_L2

Papillomaviruses are encapsulated double stranded DNA viruses. Plants are susceptible to infection by double stranded DNA viruses such as Cauliflower Mosaic virus (CaMV). The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of CaMV is thought to be involved in intra- and inter-cellular movement within the plant [1]. Engineering of proteins having similarity to papillomavirus coat proteins may enable the production of plants having better resistance to natural plant double stranded DNA viruses.

Refs

1 Thompson SR, Melcher U (1993) J Gen Virol 74: 1141-8.

Ad. Peptidase_M41

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are integral membrane proteins. They seem to be involved in the degradation of carboxy-

terminal-tagged cytoplasmic proteins. In plants, these proteins are located in the thylakoid membranes of the chloroplasts, their expression is light regulated and they are thought to be involved in degradation of soluble stromal proteins and turn-over of thylkoid proteins [1]. Manipulation of expression and structure of these proteins would have effects on the efficiency of photosynthesis and the development of chloroplasts.

Refs

1 Lindahl M, Tabak s, Cseke L, Pichersky E, Andersson B, Adam Z (1996) J Biol Chem 271: 29329-34.

Ae. UPF0051

There is some evidence that, in plants, proteins in this family are involved in ATP synthesis in chloroplasts [1, 2]. Mutations in these proteins or altering their expression would affect the efficiency of photosynthesis and energy production.

Refs

1 Kostrzewa M, Zetsche K (1992) J Mol Biol 227: 961-70.

2 Kostrzewa M, Zetsche K (1993) Plant Mol Biol 23: 67-76

Af. E7

Papillomaviruses are encapsulated double stranded DNA viruses. The Papillomavirus early protein 7 (E7) is known as a potent immortalizing and transforming agent. Transformation by E7 is thought to be mediated by the physical association of E7 with cellular proteins regulating entry into the cell cycle [1]. The result is entry into the cell cycle and suppression of terminal differentiation in mammalian cells. Thus, engineering of proteins having similarity to papillomavirus E7 protein enables the production of plants having altered cellular proliferation characteristics and possibly altered morphology. For example, overexpression of E7-like proteins would be expected to result in proliferation of cells of the tissue in which the E7 protein is expressed, perhaps with suppression of differentiation events. Thus, for example, overexpression of E7-like proteins in meristem cells can result in taller plants and suppression of leafing and/or flowering.

Refs

1 Zwerschke W, Jansen-Durr P Adv Cancer Res 2000;78:1-29

Ag. Peptidase U7

This protein is known to be an integral membrane protein in the cyanobacterium Synechocystis where it functions to digest cleaved signal peptides [1]. This activity is necessary to maintain proper secretion of mature proteins across the membrane. In higher plants this protein may be present in the plastid or chloroplast membranes where it would function by enabling protein movement into and out of the chloroplasts. Mutations in this protein would be expected to affect the development of plastids, including chloroplasts, or alter the energy transfer system within the chloroplasts, thereby affecting growth and development.

Refs

1 Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) DNA Res 3:109-36.

Ah. 5'-3' Exonuclease

The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence:

IMKKKLLLVDGSSLAFRAFFALPPLTNSAGEPTNAVYGFLKMLIKLIEQEQPTHIAVV
FDAKAKTFRHELYEGYKAGRAP
TPDELREQIPLIKELLDALGIPLLEVAGYEADDVIGTLAKLAEKEGYEVLIVTGDRDLL
QLVSDHVTVIITKKGIAEFTL
FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLLQEYGSLEGIYANL
DKLKGKKLREKLLAHKEDAKL
SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE

Ref:

Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997).

Tishkoff et al. Cancer Res. 0:0-0(1998).

Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).

Table A

Pfam	Prosite	Full Name	Description
3_5_exonuclease		3'-5' exonuclease	Accession number: PF01612 Definition: 3'-5' exonuclease Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_659 (release 4.1) Gathering cutoffs: -11 -11 Trusted cutoffs: -10.70 -10.70 Noise cutoffs: -24.50 -24.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 85137890 Reference Title: Structure of large fragment of Escherichia coli DNA Reference Title: polymerase I complexed with dTMP. Reference Author: Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA; Reference Location: Nature 1985;313:762-766. Reference Number: [2] Reference Medline: 98060913 Reference Title: The proofreading domain of Escherichia coli DNA polymerase Reference Title: I and other DNA and/or RNA exonuclease domains. Reference Author: Moser MJ, Holley WR, Chatterjee A, Mian IS; Reference Location: Nucleic Acids Res 1997;25:5110-5118. Reference Number: [3] Reference Medline: 98361165 Reference Title: Replication focus-forming activity 1 and the Werner Reference Title: syndrome gene product Reference Author: Yan H, Chen CY, Kobayashi R, Newport J; Reference Location: Nat Genet 1998;19:375-378. Reference Number: [4] Reference Medline: 97434221 Reference Title: The Werner syndrome protein is a DNA helicase. Reference Author: Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL, Reference Author: Martin GM, Oshima J, Loeb LA; Reference Location: Nat Genet 1997;17:100-103. Reference Number: [5] Reference Medline: 97370026 Reference Title: DNA helicase activity in Werner's syndrome gene product Reference Title: synthesized in a baculovirus system. Reference Author: Suzuki N, Shimamoto A, Imamura O, Kuromitsu J, Kitao S, Reference Author: Goto M, Furuichi Y; Reference Location: Nucleic Acids Res 1997;25:2973-2978. Database Reference: SCOP; 1dpi; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002562; Database Reference PDB; 1kfd ; 348; 518; Database Reference PDB; 1d8y A; 348; 518; Database Reference PDB; 1d9d A; 348; 518; Database Reference PDB; 1d9f A; 348; 518; Database Reference PDB; 1kfs A; 348; 518; Database Reference PDB; 1kln A; 348; 518; Database Reference PDB; 1krp A; 348; 518; Database Reference PDB; 1ksp A; 348; 518; Database Reference PDB; 1qsl A; 348; 518; Database Reference PDB; 2kfn A; 348; 518; Database Reference PDB; 2kfz A; 348; 518; Database Reference PDB; 2kzm A; 348; 518; Database Reference PDB; 2kzz A; 348; 518; Comment: This domain is responsible for the 3'-5' exonuclease proofreading Comment: activity of E. coli DNA polymerase I (polI) and other enzymes,

Pfam	Prosite	Full Name	Description
			Comment: it catalyses the hydrolysis of unpaired or mismatched nucleotides. Comment: This domain consists of the amino-terminal half of the Klenow fragment Comment: in E. coli poll it is also found in the Werner syndrome helicase Comment: (WRN), focus forming activity 1 protein (FFA-1) and ribonuclease D Comment: (RNase D). Comment: Werner syndrome is a human genetic disorder causing premature aging; Comment: the WRN protein has helicase activity in the 3'-5' direction [4,5]. Comment: The FFA-1 protein is required for formation of a replication foci Comment: and also has helicase activity; it is a homologue of the WRN Comment: protein [3]. Comment: RNase D is a 3'-5' exonuclease involved in tRNA processing. Comment: Also found in this family is the autoantigen PM/Scl thought to be Comment: involved in polymyositis-scleroderma overlap syndrome. Number of members: 41
3HCDH	PDOC00065	3-hydroxyacyl-CoA dehydrogenase signature	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) (HCDH) [1] is an enzyme involved in fatty acid metabolism, it catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. Most eukaryotic cells have 2 fatty-acid beta-oxidation systems, one located in mitochondria and the other in peroxisomes. In peroxisomes 3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA hydratase (ECH) and 3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme where the N- terminal domain bears the hydratase/isomerase activities and the C-terminal domain the dehydrogenase activity. There are two mitochondrial enzymes: one which is monofunctional and the other which is, like its peroxisomal counterpart, multifunctional. In Escherichia coli (gene fadB) and Pseudomonas fragi (gene faoA) HCDH is part of a multifunctional enzyme which also contains an ECH/ECI domain as well as a 3-hydroxybutyryl-CoA epimerase domain [2]. The other proteins structurally related to HCDH are: - Bacterial 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) which reduces 3-hydroxybutanoyl-CoA to acetoacetyl-CoA [3]. - Eye lens protein lambda-crystallin [4], which is specific to lagomorphes (such as rabbit). There are two major region of similarities in the sequences of proteins of the HCDH family, the first one located in the N-terminal, corresponds to the NAD- binding site, the second one is located in the center of the sequence. We have chosen to derive a signature pattern from this central region. Description of pattern(s) and/or profile(s) Consensus pattern [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-

Pfam	Prosite	Full Name	Description
			[PA]-[LIVMFY](2)- x(5)-[LIVMFYCT]-[LIVMFY]-x(2)-[GV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Pattern and text revised. References [1] Birktoff J.J., Holden H.M., Hamlin R., Xuong N.-H., Banaszak L.J. Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987). [2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:4937-4937(1990). [3] Mullany P., Clayton C.L., Pallen M.J., Slone R., Al-Saleh A., Tabaqchali S. FEMS Microbiol. Lett. 124:61-67(1994). [4] Mulders J.W.M., Hendriks W., Blankesteijn W.M., Bloemendal H., de Jong W.W. J. Biol. Chem. 263:15462-15466(1988).
4HPPD_C		4-hydroxyphenylpyruvate dioxygenase C terminal domain	Accession number: PF01626 Definition: 4-hydroxyphenylpyruvate dioxygenase C terminal domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1116 (release 4.1) Gathering cutoffs: -35 -35 Trusted cutoffs: -25.80 -25.80 Noise cutoffs: -44.90 -44.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93279307 Reference Title: Human 4-hydroxyphenylpyruvate dioxygenase. Primary Reference Title: structure and chromosomal localization of the gene. Reference Author: Ruetschi U, Dellsen A, Sahlin P, Stenman G, Rymo L, Reference Author: Lindstedt S; Reference Location: Eur J Biochem 1993;213:1081-1089. Database Reference INTERPRO; IPR002887; Comment: 4-Hydroxyphenylpyruvic acid dioxygenase (HPD) is an important enzyme Comment: in tyrosine catabolism in most organisms. A genetic deficiency in Comment: this enzyme in humans and mice leads to hereditary tyrosinemia type 3. Comment: The identity of the C-terminus of the HPD makes this part of the Comment: molecule a candidate for a functional role in the catalytic process Comment: [1]. This region is found as a separate protein Swiss:Q49717 that Comment: is somewhat different from HPD and may have a different but related Comment: protein function (Unpublished observation Bateman A). Number of members: 28

Pfam	Prosite	Full Name	Description
5_3_exonuclease		5'-3' exonuclease domain	The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence: IMKKKLLLVDGSSLAFRAFFALPPLTNSAGEPTNAVYGFLKMLIK LIEQEQPTHIAVVFDAKAKTFRHELYEGYKAGRAP TPDELREQIPLIKELLDALGIPLLEVAGYEADDVIGTLAKLAEKEG YEVLIVTGDRDLLQLVSDHVTVIITKKGIAEFTL FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLL QEYGSLEGIYANLDKLKGKKLREKLLAHKEDAKL SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE Ref: Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997). Tishkoff et al. Cancer Res. 0:0-0(1998). Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).
60s_ribosomal		60s Acidic ribosomal protein	Accession number: PF00428 Definition: 60s Acidic ribosomal protein Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_151 (release 1.0) Gathering cutoffs: 17 17 Trusted cutoffs: 17.80 17.80 Noise cutoffs: 9.30 9.30 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96282699 Reference Title: Proteins P1, P2, and P0, components of the eukaryotic Reference Title: ribosome stalk. New structural and functional aspects. Reference Author: Remacha M, Jimenez-Diaz A, Santos C, Briones E, Zambrano R, Reference Author: Rodriguez Gabriel MA, Guarinos E, Ballesta JP; Reference Location: Biochem Cell Biol 1995;73:959-968. Database Reference INTERPRO; IPR001813; Database reference: PFAMB; PB002218; Comment: This family includes archaebacterial L12, eukaryotic P0, P1 and P2. Number of members: 109
6PF2K	PDOC00158	Phosphoglycerate mutase family phosphohistidine signature	Phosphoglycerate mutase (EC 5.4.2.1) (PGAM) and bisphosphoglycerate mutase (EC 5.4.2.4) (BPGM) are structurally related enzymes which catalyze reactions involving the transfer of phospho groups between the three carbon atoms of phosphoglycerate [1,2]. Both enzymes can catalyze three different reactions, although in different proportions: - The isomerization of 2-phosphoglycerate (2-PGA) to 3-phosphoglycerate (3- PGA) with 2,3-diphosphoglycerate (2,3-DPG) as the primer of the reaction. - The synthesis of 2,3-DPG from 1,3-DPG with 3-PGA as a primer. - The degradation of 2,3-DPG to 3-PGA (phosphatase EC 3.1.3.13 activity). In mammals, PGAM is a dimeric protein. There are two isoforms of PGAM: the M

Pfam	Prosite	Full Name	Description
			(muscle) and B (brain) forms. In yeast, PGAM is a tetrameric protein. BPGM is a dimeric protein and is found mainly in erythrocytes where it plays a major role in regulating hemoglobin oxygen affinity as a consequence of controlling 2,3-DPG concentration. The catalytic mechanism of both PGAM and BPGM involves the formation of a phosphohistidine intermediate [3]. The bifunctional enzyme 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase (EC 2.7.1.105 and EC 3.1.3.46) (PF2K) [4] catalyzes both the synthesis and the degradation of fructose-2,6-bisphosphate. PF2K is an important enzyme in the regulation of hepatic carbohydrate metabolism. Like PGAM/BPGM, the fructose-2,6-bisphosphatase reaction involves a phosphohistidine intermediate and the phosphatase domain of PF2K is structurally related to PGAM/BPGM. The bacterial enzyme alpha-ribazole-5'-phosphate phosphatase (gene cobC) which is involved in cobalamin biosynthesis also belongs to this family [5]. We built a signature pattern around the phosphohistidine residue. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-x-R-H-G-[EQ]-x(3)-N [H is the phosphohistidine residue] Sequences known to belong to this class detected by the pattern ALL, except for Haemophilus influenzae PGAM. Other sequence(s) detected in SWISS-PROT 2. Note some organisms harbor a form of PGAM independent of 2,3-DPG, this enzyme is not related to the family described above [6]. Last update November 1995 / Text revised. References [1] Le Boulch P., Joulin V., Garel M.-C., Rosa J., Cohen-Solal M. Biochem. Biophys. Res. Commun. 156:874-881(1988). [2] White M.F., Fothergill-Gilmore L.A. FEBS Lett. 229:383-387(1988). [3] Rose Z.B. Meth. Enzymol. 87:43-51(1982). [4] Bazan J.F., Fletterick R.J., Pilkis S.J. Proc. Natl. Acad. Sci. U.S.A. 86:9642-9646(1989). [5] O'Toole G.A., Trzebiatowski J.R., Escalante-Semerena J.C. J. Biol. Chem. 269:26503-26511(1994). [6] Grana X., De Lecea L., El-Maghrabi M.R., Urena J.M., Caellas C., Carreras J., Puigdomenech P., Pilkis S.J., Climent F. J. Biol. Chem. 267:12797-12803(1992).
7tm_5		7TM chemoreceptor	Accession number: PF01604 Definition: 7TM chemoreceptor

Pfam	Prosite	Full Name	Description
			Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_942 (release 4.1) Gathering cutoffs: -46 -46 Trusted cutoffs: -44.30 -44.30 Noise cutoffs: -47.80 -47.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98248686 Reference Title: Two large families of chemoreceptor genes in the nematodes Reference Title: Caenorhabditis elegans and Caenorhabditis briggsae reveal Reference Title: extensive gene duplication, diversification, movement, and Reference Title: intron loss. Reference Author: Robertson HM; Reference Location: Genome Res 1998;8:449-463. Database Reference INTERPRO; IPR003003; Comment: This large family of proteins are related to 7tm_1. Comment: They are 7 transmembrane receptors. This family does not Comment: include all known members, as there are problems with Comment: overlapping specificity with 7tm_1. Comment: This family is greatly expanded in the nematode worm C. Comment: elegans. Number of members: 180
Aa_trans		Transmembrane amino acid transporter protein	Accession number: PF01490 Definition: Transmembrane amino acid transporter protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_419 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 150.80 150.80 Noise cutoffs: 3.60 3.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98007977 Reference Title: Identification and characterization of the vesicular GABA Reference Title: transporter. Reference Author: McIntire SL, Reimer RJ, Schuske K, Edwards RH, Jorgensen Reference Author: EM; Reference Location: Nature 1997;389:870-876. Database Reference INTERPRO; IPR002422; Database reference: PFAMB; PB020912; Comment: This transmembrane region is found in many amino acid transporters Comment: including UNC-47 and MTR. UNC-47 encodes a vesicular amino butyric acid Comment: (GABA) transporter, (VGAT). UNC-47 is predicted to have 10 transmembrane Comment: domains Swiss:P34579 [1]. MTR is a N system amino acid transporter system Comment: protein involved in methyltryptophan resistance Swiss:P38680. Comment: Other members of this family include proline transporters and amino Comment: acid permeases. Number of members: 50
ABC_tran	PDOC00185	ABC transporters family signature	On the basis of sequence similarities a family of related ATP-binding proteins has been characterized [1 to 5]. These proteins are associated with a variety of distinct biological processes in both prokaryotes and

Pfam	Prosite	Full Name	Description
			eukaryotes, but a majority of them are involved in active transport of small hydrophilic molecules across the cytoplasmic membrane. All these proteins share a conserved domain of some two hundred amino acid residues, which includes an ATP-binding site. These proteins are collectively known as ABC transporters. Proteins known to belong to this family are listed below (references are only provided for recently determined sequences). In prokaryotes: - Active transport systems components: alkylphosphonate uptake(phnC/phnK/ phnL); arabinose (araG); arginine (artP); dipeptide (dciAD;dppD/dppF); ferric enterobactin (fepC); ferrichrome (fhuC); galactoside (mglA); glutamine (glnQ); glycerol-3-phosphate (ugpC); glycine betaine/L-proline (proV); glutamate/aspatate (gltL); histidine (hisP); iron(III) (sfuC), iron(III) dicitrate (fecE); lactose (lacK); leucine/isoleucine/valine (braF/braG;livF/livG); maltose (malK); molybdenum (modC); nickel (nikD/ nikE); oligopeptide (amiE/amiF;oppD/oppF); peptide (sapD/sapF); phosphate (pstB); putrescine (potG); ribose (rbsA); spermidine/putrescine (potA); sulfate (cysA); vitamin B12 (btuD). - Hemolysin/leukotoxin export proteins hlyB, cyaB and lktB. - Colicin V export protein cvaB. - Lactococcin export protein lcnC [6]. - Lantibiotic transport proteins nisT (nisin) and spaT (subtilin). - Extracellular proteases B and C export protein prtD. - Alkaline protease secretion protein aprD. - Beta-(1,2)-glucan export proteins chvA and ndvA. - Haemophilus influenzae capsule-polysaccharide export protein bexA. - Cytochrome c biogenesis proteins ccmA (also known as cycV and helA). - Polysialic acid transport protein kpsT. - Cell division associated ftsE protein (function unknown). - Copper processing protein nosF from Pseudomonas stutzeri. - Nodulation protein nodI from Rhizobium (function unknown). - Escherichia coli proteins cydC and cydD. - Subunit A of the ABC excision nuclease (gene uvrA). - Erythromycin resistance protein from Staphylococcus epidermidis (gene msrA). - Tylosin resistance protein from Streptomyces fradiae (gene tlrC) [7]. - Heterocyst differentiation protein (gene hetA) from Anabaena PCC 7120. - Protein P29 from Mycoplasma hyorhinis, a probable component of a high affinity transport system. - yhbG, a putative protein whose gene is linked with ntrA in many bacteria such as Escherichia coli, Klebsiella pneumoniae, Pseudomonas putida, Rhizobium meliloti and Thiobacillus ferrooxidans. - Escherichia coli and related bacteria hypothetical proteins yabJ, yadG, yagC, ybbA, ycjW, yddA, yehX, yejF, yheS, yhiG, yhiH, yjcW, yjjK, yojI, yrbF and ytfR. In eukaryotes:

Pfam	Prosite	Full Name	Description
			- The multidrug transporters (Mdr) (P-glycoprotein), a family of closely related proteins which extrude a wide variety of drugs out of the cell (for a review see [8]). - Cystic fibrosis transmembrane conductance regulator (CFTR), which is most probably involved in the transport of chloride ions. - Antigen peptide transporters 1 (TAP1, PSF1, RING4, HAM-1, mtp1) and 2 (TAP2, PSF2, RING11, HAM-2, mtp2), which are involved in the transport of antigens from the cytoplasm to a membrane-bound compartment for association with MHC class I molecules. - 70 Kd peroxisomal membrane protein (PMP70). - ALDP, a peroxisomal protein involved in X-linked adrenoleukodystrophy [9]. - Sulfonylurea receptor [10], a putative subunit of the B-cell ATP-sensitive potassium channel. - Drosophila proteins white (w) and brown (bw), which are involved in the import of ommatidium screening pigments. - Fungal elongation factor 3 (EF-3). - Yeast STE6 which is responsible for the export of the a-factor pheromone. - Yeast mitochondrial transporter ATM1. - Yeast MDL1 and MDL2. - Yeast SNQ2. - Yeast sporidesmin resistance protein (gene PDR5 or STS1 or YDR1). - Fission yeast heavy metal tolerance protein hmt1. This protein is probably involved in the transport of metal-bound phytochelatins. - Fission yeast brefeldin A resistance protein (gene bfr1 or hba2). - Fission yeast leptomycin B resistance protein (gene pmd1). - mbpX, a hypothetical chloroplast protein from Liverwort. - Prestalk-specific protein tagB from slime mold. This protein consists of two domains: a N-terminal subtilase catalytic domain (see <PDOC00125>) and a C-terminal ABC transporter domain. As a signature pattern for this class of proteins, we use a conserved region which is located between the 'A' and the 'B' motifs of the ATP-binding site. **Consensus pattern** [LIVMFYC]-[SA]-[SAPGLVFYKQH]-G-[DENQMW]-[KRQASPCLIMFW]- [KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]-{PHY}-[LIVMFW]- [SAGCLIVP]-{FYWHP}-{KRHP}-[LIVMFYWSTA] **Sequences known to belong to this class detected by the pattern** ALL, except for 25 sequences. **Other sequence(s) detected in SWISS-PROT** 42. **Note** the ATP-binding region is duplicated in araG, mdl, msrA, rbsA, tlrC, uvrA, yejF, Mdr's, CFTR, pmd1 and in EF-3. In some of those proteins, the above pattern only detect one of the two copies of the domain. **Note** the proteins belonging to this family also contain one or two copies of the ATP-binding motifs 'A' and 'B' (see <PDOC00017>). July 1998 / Text revised. [1] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [2] Higgins C.F., Gallagher M.P., Mimmack M.M., Pearce S.R. BioEssays 8:111-116(1988). [3] Higgins C.F., Hiles I.D., Salmond G.P.C., Gill D.R., Downie J.A., Evans I.J., Holland I.B., Gray L., Buckels S.D., Bell A.W., Hermodson M.A. Nature 323:448-450(1986).

Pfam	Prosite	Full Name	Description
			[4] Doolittle R.F., Johnson M.S , Husain I., van Houten B., Thomas D.C., Sancar A. Nature 323:451-453(1986). [5] Blight M.A., Holland I.B. Mol. Microbiol. 4:873-880(1990). [6] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L. Appl. Environ. Microbiol. 58:1952-1961(1992). [7] Rosteck P.R. Jr., Reynolds P.A., Hershberger C.L. Gene 102:27-32(1991). [8] Gottesman M.M., Pastan I. J. Biol. Chem. 263:12163-12166(1988). [9] Valle D., Gaertner J. Nature 361:682-683(1993). [10] Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV, Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J., Nelson D.A. Science 268:423-426(1995).
ABC2_membrane	PDOC00692	ABC-2 type transport system integral membrane proteins signature	Integral membrane components of a number of bacterial active transport systems have been shown to be evolutionary related and to form a distinct family [1,2]. These proteins are: - Escherichia coli kpsM, involved in polysialic acid export. - Haemophilus influenzae bexB, involved in polyribosylribitol phosphate capsule polysaccharide export. - Salmonella typhi vexB, involved in translocation of the Vi polysaccharide. - Neisseria meningitidis ctrC, involved in polyneuraminic acid capsule polysaccharide export. - Rhizobiacae nodulation protein J (gene nodJ), probably involved in exporting a modified beta-1,4-linked N-acetylglucosamine oligosaccharide. - Streptomyces peucetius drrB, involved in exporting the antibiotics daunorubicin and doxorubicin. - Klebsiella pneumoniae O-antigen exprt system protein rfbA. - Yersinia enterocolitica O-antigen exprt system protein rfbD. - Escherichia coli hypothetical protein yadH. - Escherichia coli hypothetical protein yhhJ. The molecular size of these proteins is around 30 Kd. They are thought to contain six transmembrane regions. They either form homooligomeric channels or associate with another type of transmembrane protein to form heteroligomers. Transport systems in which they participate are energized by an ATP-binding protein that belongs to the ABC transporter family. The designation 'ABC-2' has been proposed [1] for these transport systems. As a signature pattern, we selected a conserved region located in the C-terminal section of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)- [LIMGA]-[PGSNQ]-x(9,12)-P-[LIMFT]-x-[HRSY]-

Pfam	Prosite	Full Name	Description
			x(5)-[RQ] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update November 1997 / Pattern and text revised. References [1] Reizer J., Reizer A., Saier M.H. Jr. Protein Sci. 1:1326-1332(1992). [2] Vazquez M., Santana O., Quinto C. Mol. Microbiol. 8:369-377(1993).
ABC-3		ABC 3 transport family	Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein. References describing these family members and their biological activities: Abbot et al., J. Biol. Chem. 267:10759-10763(1992);Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum. Mol. Genet. 2:1437-1441(1993); Clifford et al., Genetics 137:531-550(1994); Cocozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al.. Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Proc. Natl. Acad. Sci. U.S A. 84:704-708(1987); Ehsani et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997). Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 266:5260-5267(1991); Lai et al., Neuron 6:691-704(1991); Lax et al., Mol. Cell. Biol. 8:1970-1978(1988); Lebrun et al., J. Biol. Chem. 268:11272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985). Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Mol. Endocrinol. 4:1183-1191(1990); Odawara et al., Science 245:66-68(1989); Raz et al., Genetics 129:191-201(1991). Sakai et al., J. Mol. Biol. 256:548-555(1996); Schaeffer et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Cell 46:1091-1101(1986); Seino et al., Biochem. Biophys. Res. Commun. 159:312-316(1989); Seino et al., Diabetes 39:123-128(1990); Semba et al., Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985); Shier et al., J. Biol. Chem. 264:14605-14608(1989); Taira et al., Science 245:63-66(1989); Tewari et al., J. Biol.

Pfam	Prosite	Full Name	Description
			Chem. 264:16238-16245(1989); Ullrich et al., Nature 313:756-761(1985). Ullrich et al., EMBO J. 5:2503-2512(1986); van der Vorm et al., Diabetologia 36:172-174(1993); van der Vorm et al., J. Biol. Chem. 267:66-71(1992); Wadsworth et al., Nature 314:178-180(1985); White et al., Cell 54:641-649(1988); Xu et al., J. Biol. Chem. 265:18673-18681(1990); Yamamoto et al., Nature 319:230-234(1986); and Yoshimasa et al., Science 240:784-787(1988).
ACAT		Sterol O-acyltransferase	Accession number: PF01800 Definition: Sterol O-acyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1454 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 112.80 112.80 Noise cutoffs: -128.10 -128.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98434592 Reference Title: Characterization of two human genes encoding acyl coenzyme Reference Title: A:cholesterol acyltransferase-related enzymes. Reference Author: Oelkers P, Behari A, Cromley D, Billheimer JT, Sturley SL; Reference Location: J Biol Chem 1998;273:26765-26771. Reference Number: [2] Reference Medline: 98434590 Reference Title: Identification of a form of acyl-CoA:cholesterol Reference Title: acyltransferase specific to liver and intestine in nonhuman Reference Title: primates. Reference Author: Anderson RA, Joyce C, Davis M, Reagan JW, Clark M, Shelness Reference Author: GS, Rudel LL; Reference Location: J Biol Chem 1998;273:26747-26754. Reference Number: [3] Reference Medline: 96243137 Reference Title: Sterol esterification in yeast: a two-gene process. Reference Author: Yang H, Bard M, Bruner DA, Gleeson A, Deckelbaum RJ, Reference Author: Aljinovic G, Pohl TM, Rothstein R, Sturley SL; Reference Location: Science 1996;272:1353-1356. Database Reference INTERPRO; IPR002688, Comment: Sterol O-acyltransferases or acyl-coa:cholesterol acyltransferase Comment: (ACAT) EC:2.3.1.26 is a transmembrane protein that catalyses the Comment: esterification of cholesterol to its cholesterol ester storage Comment: form. Number of members: 21
ACPS		4'-phosphopantetheinyl transferase superfamily	Accession number: PF01648 Definition: 4'-phosphopantetheinyl transferase superfamily Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1679 (release 4.1) Gathering cutoffs: 0 0 Trusted cutoffs: 0.60 0.60 Noise cutoffs: -4.00 -4.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96027548 Reference Title: Cloning, overproduction, and

Pfam	Prosite	Full Name	Description
			characterization of the Reference Title: Escherichia coli holo-acyl carrier protein synthase. Reference Author: Lambalot RH, Walsh CT; Reference Location: J Biol Chem 1995;270:24658-24661. Reference Number: [2] Reference Medline: 97144264 Reference Title: A new enzyme superfamily - the phosphopantetheinyl Reference Title: transferases. Reference Author: Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Reference Author: Marahiel MA, Reid R, Khosla C, Walsh CT; Reference Location: Chem Biol 1996;3:923-936. Reference Number: [3] Reference Medline: 10581256 Reference Title: Crystal structure of the surfactin synthetase-activating Reference Title: enzyme sfp. a prototype of the 4'-phosphopantetheinyl Reference Title: transferase superfamily [In Process Citation] Reference Author: Reuter K, Mofid MR, Marahiel MA, Ficner R; Reference Location: EMBO J 1999;18:6823-6831. Database Reference INTERPRO; IPR002582; Database reference: PFAMB; PB007908; Database reference: PFAMB; PB041384; Comment: Members of this family transfers the Comment: 4'-phosphopantetheine (4'-PP) moiety from coenzyme A (CoA) to Comment: the invariant serine of pp-binding. This post-translational Comment: modification renders holo-ACP capable of acyl group activation Comment: via thioesterification of the cysteamine thiol of 4'-PP [1]. Comment: This superfamily consists of two subtypes: The ACPS type Comment: such as Swiss:P24224 and the Sfp type such as Swiss:P39135. Comment: The structure of the Sfp type is known [3], which shows the Comment: active site accommodates a magnesium ion. The most highly Comment: conserved regions of the alignment are involved in binding Comment: the magnesium ion. Number of members: 46
ACT		ACT domain	Accession number: PF01842 Definition: ACT domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 25 0 Trusted cutoffs: 26.10 0.50 Noise cutoffs: 24.50 24.50 HMM build command line· hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95236205 Reference Title: The allosteric ligand site in the Vmax-type cooperative Reference Title: enzyme phosphoglycerate dehydrogenase. Reference Author: Schuller DJ, Grant GA, Banaszak LJ; Reference Location: Nat Struct Biol 1995;2·69-76. Reference Number: [2] Reference Medline: 99241053 Reference Title: Gleaning non-trivial structural, functional and Reference Title: evolutionary information about proteins by iterative

Pfam	Prosite	Full Name	Description
			Reference Title: database searches. Reference Author: Aravind L, Koonin EV; Reference Location: J Mol Biol 1999;287:1023-1040. Database Reference: SCOP; 1psd; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002912; Database Reference PDB; 1phz A; 35; 110; Database Reference PDB; 2phm A; 35; 110; Database Reference PDB; 1psd A; 338; 410; Database Reference PDB; 1psd B; 338; 410; Database reference: PFAMB; PB001977; Database reference: PFAMB; PB008097; Database reference: PFAMB; PB010480; Database reference: PFAMB; PB011031; Database reference: PFAMB; PB031880; Database reference: PFAMB; PB038464; Database reference: PFAMB; PB040963; Database reference: PFAMB; PB041518; Database reference: PFAMB; PB041667; Comment: This family of domains generally have a regulatory role. Comment: ACT domains are linked to a wide range of metabolic Comment: enzymes that are regulated by amino acid concentration. Comment: Pairs of ACT domains bind specifically to a particular Comment: amino acid leading to regulation of the linked enzyme. Comment: The ACT domain is found in: Comment: D-3-phosphoglycerate dehydrogenase EC:1.1.1.95 Swiss:P08328, Comment: which is inhibited by serine [1]. Comment: Aspartokinase EC:2.7.2.4 Swiss:P53553, which is regulated by lysine. Comment: Acetolactate synthase small regulatory subunit Swiss:P00894, Comment: which is inhibited by valine. Comment: Phenylalanine-4-hydroxylase EC:1.14.16.1 Swiss:P00439, which Comment: is regulated by phenylalanine. Comment: Prephenate dehydrogenase EC:4.2.1.51 Swiss:P21203. Comment: formyltetrahydrofolate deformylase EC:3.5.1.10, Swiss:P37051, Comment: which is activated by methionine and inhibited by glycine. Comment: GTP pyrophosphokinase EC:2.7 6.5 Swiss:P11585. Number of members: 177
Activin_recp		Activin types I and II receptor domain	Accession number: PF01064 Definition: Activin types I and II receptor domain Author: Finn RD, Bateman A Alignment method of seed: Clustalw_manual Source of seed members: Pfam-B_338 (release 3.0) Gathering cutoffs: 22 22 Trusted cutoffs: 23.10 23.10 Noise cutoffs: 11.30 21.20 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97454714 Reference Title: From receptor to nucleus: the Smad pathway. Reference Author: Baker JC, Harland RM; Reference Location: Curr Opin Genet Dev 1997;7:467-473. Reference Number: [2] Reference Medline: 94131268 Reference Title: The TGF-beta superfamily: new members, new receptors, and Reference Title: new genetic tests of function in different organisms. Reference Author: Kingsley DM;

Pfam	Prosite	Full Name	Description
			Reference Location: Genes Dev 1994;8:133-146. Reference Number: [3] Reference Medline: 93390967 Reference Title: Activin receptor-like kinases: a novel subclass of Reference Title: cell-surface receptors with predicted serine/threonine Reference Title: kinase activity. Reference Author: ten Dijke P, Ichijo H, Franzen P, Schulz P, Saras J, Reference Author: Toyoshima H, Heldin CH, Miyazono K; Reference Location: Oncogene 1993;8:2879-2887. Database Reference INTERPRO; IPR000472; Database reference: PFAMB; PB024112; Database reference: PFAMB; PB040755; Comment: This Pfam entry consists of both TGF-beta receptor types. Comment: This is an alignment of the hydrophilic cysteine-rich Comment: ligand-binding domains, Comment: Both receptor types, (type I and II) posses a 9 amino Comment: acid cysteine box, with the the consensus CCX{4-5}CN. Comment: The type I receptors also possess 7 extracellular residues Comment: preceding the cysteine box. Number of members: 79
Acyl-ACP_TE		Acyl-ACP thioesterase	Accession number: PF01643 Definition: Acyl-ACP thioesterase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_928 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 91.70 91.70 Noise cutoffs: -192.80 -192.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96068671 Reference Title: Modification of the substrate specificity of an acyl-acyl Reference Title: carrier protein thioesterase by protein engineering. Reference Author: Yuan L, Voelker TA, Hawkins DJ; Reference Location: Proc Natl Acad Sci U S A 1995;92:10639-10643. Reference Number: [2] Reference Medline: 92320297 Reference Title: Fatty acid biosynthesis redirected to medium chains in Reference Title: transgenic oilseed plants. Reference Author: Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Reference Author: Hawkins DJ, Radke SE, Davies HM; Reference Location: Science 1992;257:72-74. Database Reference INTERPRO; IPR002864; Comment: This family consists of various acyl-acyl carrier protein (ACP) Comment: thioesterases (TE) these terminate fatty acyl group extension via Comment: hydrolyzing an acyl group on a fatty acid [1]. Number of members: 30
Acyltransferase		Acyltransferase	Accession number: PF01553 Definition: Acyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_128 (release 4.0) Gathering cutoffs: 8 8 Trusted cutoffs: 14.40 14.40 Noise cutoffs: 2.50 2.50

Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97411131 Reference Title: Barth syndrome may be due to an acyltransferase deficiency. Reference Author: Neuwald AF; Reference Location: Curr Biol 1997;7:465-466. Reference Number: [2] Reference Medline: 96224398 Reference Title: A novel X-linked gene, G4.5. is responsible for Barth Reference Title: syndrome. Reference Author: Bione S, D'Adamo P, Maestrini E, Gedeon AK, Bolhuis PA, Reference Author: Toniolo D; Reference Location: Nat Genet 1996;12:385-389. Database Reference INTERPRO; IPR002123; Database reference: PFAMB; PB009622; Database reference: PFAMB; PB009717; Database reference: PFAMB; PB033259; Database reference: PFAMB; PB041102; Database reference: PFAMB; PB041638; Comment: This family contains acyltransferases involved in phospholipid Comment: biosynthesis and other proteins of unknown function [1]. This Comment: family also includes tafazzin Swiss:Q16635, the Barth syndrome Comment: gene [2]. Number of members: 74
Adaptin_N		Adaptin N terminal region	Accession number: PF01602 Definition: Adaptin N terminal region Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_491 (release 4.0) Gathering cutoffs: 12 12 Trusted cutoffs: 15.50 15.50 Noise cutoffs: 9.00 9.00 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97409270 Reference Title: Linking cargo to vesicle formation: receptor tail Reference Title: interactions with coat proteins. Reference Author: Kirchhausen T, Bonifacino JS, Riezman H; Reference Location: Curr Opin Cell Biol 1997;9:488-495. Reference Number: [2] Reference Medline: 89202379 Reference Title: Structural and functional division into two domains of the Reference Title: large (100- to 115-kDa)chains of the clathrin-associated Reference Title: protein complex AP-2. Reference Author: RAKirchhausen T, Nathanson KL, Matsui W, Vaisberg A, Chow Reference Author: EP, Burne C, Keen JH, Davis AE; Reference Location: Proc Natl Acad Sci U S A 1989;86:2612-2616. Database Reference INTERPRO; IPR002553; Database reference: PFAMB; PB040953; Comment: This family consists of the N terminal region of various alpha, Comment: beta and gamma subunits of the AP-1, AP-2 and AP-3 adaptor Comment: protein complexes. The adaptor protein (AP) complexes are involved in Comment: the formation of clathrin-coated pits and vesicles [1]. Comment: The N-terminal region of the various adaptor proteins (APs) is constant

Pfam	Prosite	Full Name	Description
			Comment: by comparison to the C-terminal which is variable within members of the Comment: AP-2 family[2]; and it has been proposed that this constant region Comment: interacts with another uniform component of the coated vesicles [2]. Number of members: 66
ALAD	PDOC00153	Delta-aminolevulinic acid dehydratase active site	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) [1] catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen. The enzyme is an oligomer composed of eight identical subunits. Each of the subunits binds an atom of zinc or of magnesium (in plants). A lysine has been implicated in the catalytic mechanism [2]. The sequence of the region in the vicinity of the active site residue is conserved in ALAD from various prokaryotic and eukaryotic species. Description of pattern(s) and/or profile(s) Consensus pattern G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y [K is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Li J.-M., Russell C.S., Cosloy S.D. Gene 75:177-184(1989). [2] Gibbs P.N.B., Jordan P.M. Biochem. J. 236:447-451(1986).
Aldolase	PDOC00144	KDPG and KHG aldolases active site signatures	4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) (KHG-aldolase) catalyzes the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate. Phospho-2-dehydro-3-deoxygluconate aldolase (EC 4.1.2.14) (KDPG-aldolase) catalyzes the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde 3-phosphate. These two enzymes are structurally and functionally related [1]. They are both homotrimeric proteins of approximately 220 amino-acid residues. They are class I aldolases whose catalytic mechanism involves the formation of a Schiff-base intermediate between the substrate and the epsilon-amino group of a lysine residue. In both enzymes, an arginine is required for catalytic activity. We developed two signature patterns for these enzymes. The first one contains the active site arginine and the second, the lysine involved in the Schiff- base formation. Description of pattern(s) and/or profile(s)

Pfam	Prosite	Full Name	Description
			Consensus pattern G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R [R is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Bacillus subtilis KDPG-aldolase which has Thr instead of Arg in the active site. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G [K is involved in Schiff-base formation] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Vlahos C J., Dekker E.E. J. Biol. Chem. 263:11683-11691(1988).
Alpha_L_fucos	PDOC00324	Alpha-L-fucosidase	Alpha-L-fucosidase (EC 3.2.1.51) [1] is a lysosomal enzyme responsible for hydrolyzing the alpha-1,6-linked fucose joined to the reducing-end N-acetylglucosamine of the carbohydrate moieties of glycoproteins. Deficiency of alpha-L-fucosidase results in the lysosomal storage disease fucosidosis. A cysteine residue is important for the activity of the enzyme. There is only one cysteine conserved between the sequence of mammalian alpha-L-fucosidase and that of the slime mold Dictyostelium discoideum. We have derived a pattern from the region around that conserved cysteine. Description of pattern(s) and/or profile(s) Consensus pattern P-x(2)-L-x(3)-K-W-E-x-C [C is the putative active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note these proteins belong to family 29 in the classification of glycosyl hydrolases [2,E1]. Last update November 1997 / Pattern and text revised. References [1] Fisher K.J., Aronson N.N. Jr. Biochem. J. 264:695-701(1989). [2] Henrissat B. Biochem. J. 280:309-316(1991). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt
Amino_oxidase		Flavin containing amine oxidase	Accession number: PF01593 Definition: Flavin containing amine oxidase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_606 (release 4.1) Gathering cutoffs: -110 -110 Trusted cutoffs: -110.00 -110.00 Noise cutoffs: -111.80 -111.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98258926

Pfam	Prosite	Full Name	Description
			Reference Title: Maize polyamine oxidase: primary structure from protein and Reference Title: cDNA sequencing. Reference Author: Tavladoraki P, Schinina ME, Cecconi F, Agostino SD, Manera Reference Author: F, Rea G, Mariottini P, Federico R, Angelini R; Reference Location: FEBS Lett 1998;426:62-66. Reference Number: [2] Reference Medline: 97306298 Reference Title: A key amino acid responsible for substrate selectivity of Reference Title: monoamine oxidase A and B. Reference Author: Tsugeno Y, Ito A; Reference Location: J Biol Chem 1997;272:14033-14036. Reference Number: [3] Reference Medline: 95287865 Reference Title: Cloning, sequencing and heterologous expression of the Reference Title: monoamine oxidase gene from Aspergillus niger. Reference Author: Schilling B, Lerch K; Reference Location: Mol Gen Genet 1995;247:430-438. Database Reference: SCOP; 1b37; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002937; Database Reference PDB; 1b37 A; 14; 455; Database Reference PDB; 1b5q A; 14; 455; Database Reference PDB; 1b37 B; 14; 455; Database Reference PDB; 1b37 C; 14; 455; Database Reference PDB; 1b5q B; 14; 455; Database Reference PDB; 1b5q C; 14; 455; Database reference: PFAMB; PB017518; Database reference: PFAMB; PB024839; Database reference: PFAMB; PB040747; Comment: This family consists of various amine oxidases, including maze polyamine Comment: oxidase (PAO) [1] and various flavin containing monoamine oxidases Comment: (MAO). The aligned region includes the flavin binding site of these Comment: enzymes. Comment: In vertebrates MAO plays an important role regulating the intracellular Comment: levels of amines via there oxidation; these include various Comment: neurotransmitters, neurotoxins and trace amines [2]. In lower eukaryotes Comment: such as aspergillus and in bacteria the main role of amine oxidases is Comment: to provide a source of ammonium [3]. Comment: PAOs in plants, bacteria and protozoa oxidase spermidine and spermine Comment: to an aminobutyral, diaminopropane and hydrogen peroxide and are Comment: involved in the catabolism of polyamines [1]. Comment: Other members of this family include tryptophan 2-monooxygenase, Comment: putrescine oxidase, corticosteroid binding proteins and antibacterial Comment: glycoproteins. Number of members: 58
ANF_receptor	PDOC00430	Natriuretic peptides receptors signature	Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger. Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which

Pfam	Prosite	Full Name	Description
			seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction. GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C- terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short. We developed a pattern from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N- terminal part of the domain. Description of pattern(s) and/or profile(s) Consensus pattern G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Garbers D.L. New Biol. 2:499-504(1990). [2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).
Apocytochrome_F	PDOC00169	Cytochrome c family heme-binding site signature	In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c. Description of pattern(s) and/or profile(s) Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the pattern ALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT 454.

Pfam	Prosite	Full Name	Description
			Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 ! Last update June 1992 / Text revised. References [1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).
arf	PDOC00781 PDOC00017 PDOC01020	ADP-ribosylation factors family signature; ATP/GTP-binding site motif A (P-loop); ATP phosphoribosyltransferase signature PROSITE cross-reference(s)	ADP-ribosylation factors (ARF) [1,2,3,4] are 20 Kd GTP-binding proteins involved in protein trafficking. They may modulate vesicle budding and uncoating within the Golgi apparatus. ARF's also act as allosteric activators of cholera toxin ADP-ribosyltransferase activity. They are evolutionary conserved and present in all eukaryotes. At least six forms of ARF are present in mammals and three in budding yeast. The ARF family also includes proteins highly related to ARF's but which lack the cholera toxin cofactor activity, they are collectively known as ARL's (ARF-like). ARD1 is a 64 Kd mammalian protein of unknown biological function that contains an ARF domain at its C-terminal extremity. Proteins from the ARF family are generally included in the RAS 'superfamily' of small GTP-binding proteins [5], but they are only slightly related to the other RAS proteins. They also differ from RAS proteins in that they lack cysteine residues at their C-termini and are therefore not subject to prenylation. The ARFs are N-terminally myristoylated (the ARLs have not yet been shown to be modified in such a fashion). As a signature pattern, we selected a conserved region in the C-terminal part of ARF's and ARL's. Description of pattern(s) and/or profile(s) Consensus pattern [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)- [GSA]-[LIVMF]-x-[WK]-[LIVM] Sequences known to belong to this class detected by the pattern ALL, except for 4 sequences. Other sequence(s) detected in SWISS-PROT NONE. Note proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see <PDOC00017>). Expert(s) to contact by email Kahn R.A. rkahn@bimcore.emory.edu Last update November 1997 / Pattern and text revised. Cell. Signal. 4.367-399(1993). References [1] Boman A.L., Kahn R.A. Trends Biochem. Sci. 20:147-150(1995). [2] Moss J., Vaughan M. [3] Moss J., Vaughan M.

Pfam	Prosite	Full Name	Description
			Prog. Nucleic Acid Res. Mol. Biol. 45:47-65(1993). [4] Amor J.C., Harrison D.H., Kahn R.A., Ringe D. Nature 372:704-708(1994). [5] Valencia A., Chardin P., Wittinghofer A., Sander C. Biochemistry 30:4637-4648(1991). From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5]. There are numerous ATP- or GTP-binding proteins in which the P-loop is found. We list below a number of protein families for which the relevance of the presence of such motif has been noted: - ATP synthase alpha and beta subunits (see <PDOC00137>). - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>). - Dynamins and dynamin-like proteins (see <PDOC00362>). - Guanylate kinase (see <PDOC00670>). - Thymidine kinase (see <PDOC00524>). - Thymidylate kinase (see <PDOC01034>). - Shikimate kinase (see <PDOC00868>). - Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>). - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>). - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran (see <PDOC00859>). - ADP-ribosylation factors family (see <PDOC00781>). - Bacterial dnaA protein (see <PDOC00771>). - Bacterial recA protein (see <PDOC00131>). - Bacterial recF protein (see <PDOC00539>). - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family (See <PDOC00388>). - Bacterial type II secretion system protein E (see <PDOC00567>). Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for adenylate kinase, in which there is a single deviation from the P-loop

Pfam	Prosite	Full Name	Description
			pattern: in the last position Gly is found instead of Ser or Thr. Description of pattern(s) and/or profile(s) Consensus pattern [AG]-x(4)-G-K-[ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferritin light chain). Expert(s) to contact by email Koonin E.V. koonin@ncbi.nlm.nih.gov Last update July 1999 / Text revised. References [1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982). [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985). [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986). [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987). [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990). [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993). [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990). [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata). [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989). [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989). ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM]

Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / First entry.
ArgJ		ArgJ family	Accession number: PF01960 Definition: ArgJ family Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 258.70 99.60 Noise cutoffs: 7.10 7.10 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93232760 Reference Title: Primary structure, partial purification and regulation of Reference Title: key enzymes of the acetyl cycle of arginine biosynthesis in Reference Title: Bacillus stearothermophilus: dual function of ornithine Reference Title: acetyltransferase. Reference Author: Sakanyan V, Charlier D, Legrain C, Kochikyan A, Mett I, Reference Author: Pierard A, Glansdorff N; Reference Location: J Gen Microbiol 1993;139:393-402. Database Reference INTERPRO; IPR002813; Comment: Members of the ArgJ family catalyse the first EC:2.3.1.35 and Comment: fifth steps EC:2.3.1.1 in arginine biosynthesis. Number of members: 22
Armadillo_seg		Armadillo/beta-catenin-like repeats	Accession number: PF00514 Definition: Armadillo/beta-catenin-like repeats Author: Bateman A, Chris Ponting, Joerg Schultz, Peer Bork Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs. 24 0 Trusted cutoffs: 24.10 0.00 Noise cutoffs: 20.70 20.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97442350 Reference Title: Three-dimensional structure of the armadillo repeat region Reference Title: of beta-catenin. Reference Author: Huber AH, Nelson WJ, Weis WI; Reference Location: Cell 1997;90:871-882. Reference Number: [2] Reference Medline: 96107551 Reference Title: Signal transduction of beta-catenin. Reference Author: Gumbiner BM; Reference Location: Curr Opin Cell Biol 1995;7:634-640. Reference Number: [3] Reference Medline: 97454713 Reference Title: Armadillo and dTCF: a marriage made in the nucleus. Reference Author: Cavallo R, Rubenstein D, Peifer M; Reference Location: Curr Opin Genet Dev 1997;7:459-466. Reference Number: [4] Reference Medline: 94082295 Reference Title: Association of the APC tumor suppressor protein with Reference Title: catenins. Reference Author: Su LK, Vogelstein B, Kinzler KW; Reference Location: Science 1993;262:1734-1737. Reference Number: [5] Reference Medline: 94082294

Pfam	Prosite	Full Name	Description
			Reference Title: Association of the APC gene product with beta-catenin. Reference Author: Rubinfeld B, Souza B, Albert I, Muller O, Chamberlain SH, Reference Author: Masiarz FR, Munemitsu S, Polakis P; Reference Location: Science 1993;262:1731-1734. Reference Number: [6] Reference Medline: 91084846 Reference Title: The segment polarity gene armadillo encodes a functionally Reference Title: modular protein that is the Drosophila homolog of human Reference Title: plakoglobin. Reference Author: Peifer M, Wieschaus E; Reference Location: Cell 1990;63:1167-1176. Database Reference: SCOP; 3bct; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: EXPERT; Chris.Ponting@human-anatomy.oxford.ac.uk; Database reference: SMART; ARM; Database Reference INTERPRO; IPR000225; Database Reference PDB; 1ee5 A; 417; 457; Database Reference PDB; 1bk5 A; 417; 457; Database Reference PDB; 1bk5 B; 417; 457; Database Reference PDB; 1bk6 A; 417; 457; Database Reference PDB; 1bk6 B; 417; 457; Database Reference PDB; 1ee4 A; 417; 457; Database Reference PDB; 1ee4 B; 417; 457; Database Reference PDB; 1ejl I; 409; 449; Database Reference PDB; 1ejy I; 409; 449; Database Reference PDB; 1ial A; 409; 449; Database Reference PDB; 1ee5 A; 246; 286; Database Reference PDB; 1bk5 A; 246; 286; Database Reference PDB; 1bk5 B; 246; 286; Database Reference PDB; 1bk6 A; 246; 286; Database Reference PDB; 1bk6 B; 246; 286; Database Reference PDB; 1ee4 A; 246; 286; Database Reference PDB; 1ee4 B; 246, 286; Database Reference PDB; 1ejl I; 241; 280; Database Reference PDB; 1ejy I; 241; 280; Database Reference PDB; 1ial A; 241; 280; Database Reference PDB; 1ee5 A; 288; 328; Database Reference PDB; 1bk5 A; 288; 328; Database Reference PDB; 1bk5 B; 288; 328; Database Reference PDB; 1bk6 A; 288; 328; Database Reference PDB; 1bk6 B; 288; 328; Database Reference PDB; 1ee4 A; 288; 328; Database Reference PDB; 1ee4 B; 288; 328; Database Reference PDB; 1ejl I; 282; 322; Database Reference PDB; 1ejy I; 282; 322; Database Reference PDB; 1ial A; 282; 322; Database Reference PDB; 1ejl I; 151; 191; Database Reference PDB; 1ejy I; 151; 191; Database Reference PDB; 1ial A; 151; 191; Database Reference PDB; 1ee5 A; 162; 202; Database Reference PDB; 1bk5 A; 162; 202; Database Reference PDB; 1bk5 B; 162; 202; Database Reference PDB; 1bk6 A; 162; 202; Database Reference PDB; 1bk6 B; 162; 202; Database Reference PDB; 1ee4 A; 162; 202; Database Reference PDB; 1ee4 B; 162; 202; Database Reference PDB; 1ee5 A; 330; 370; Database Reference PDB; 1bk5 A; 330; 370; Database Reference PDB; 1bk5 B; 330; 370; Database Reference PDB; 1bk6 A; 330; 370; Database Reference PDB; 1bk6 B; 330; 370; Database Reference PDB; 1ee4 A; 330; 370; Database Reference PDB; 1ee4 B; 330; 370; Database Reference PDB; 1ejl I; 324; 364; Database Reference PDB; 1ejy I; 324; 364; Database Reference PDB; 1ial A; 324; 364; Database Reference PDB; 1ee5 A; 372; 412; Database Reference PDB; 1bk5 A; 372; 412; Database Reference PDB; 1bk5 B; 372; 412;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1bk6 A; 372; 412; Database Reference PDB; 1bk6 B; 372; 412; Database Reference PDB; 1ee4 A; 372; 412; Database Reference PDB; 1ee4 B; 372; 412; Database Reference PDB; 1ejl I; 366; 406; Database Reference PDB; 1ejy I; 366; 406; Database Reference PDB; 1ial A; 366; 406; Database Reference PDB; 1ejl I; 108; 149; Database Reference PDB; 1ejy I; 108; 149; Database Reference PDB; 1ial A; 108; 149; Database Reference PDB; 1ee5 A; 119; 160; Database Reference PDB; 1bk5 A; 119; 160; Database Reference PDB; 1bk5 B; 119; 160; Database Reference PDB; 1bk6 A; 119; 160; Database Reference PDB; 1bk6 B; 119; 160; Database Reference PDB; 1ee4 A; 119; 160; Database Reference PDB; 1ee4 B; 119; 160; Database Reference PDB; 3bct ; 583; 623; Database Reference PDB; 2bct ; 583; 623; Database Reference PDB; 3bct ; 391; 429; Database Reference PDB; 2bct ; 391; 429; Database Reference PDB; 3bct ; 224; 264; Database Reference PDB; 2bct ; 224; 264; Database Reference PDB; 3bct ; 431; 473; Database Reference PDB; 2bct ; 431; 473; Database Reference PDB; 3bct ; 350; 390; Database Reference PDB; 2bct ; 350; 390; Database Reference PDB; 1ejl I; 193; 238; Database Reference PDB; 1ejy I; 193; 238; Database Reference PDB; 1ial A; 193; 238; Database Reference PDB; 1ee5 A; 204; 244; Database Reference PDB; 1bk5 A; 204; 244; Database Reference PDB; 1bk5 B; 204; 244; Database Reference PDB; 1bk6 A; 204; 244; Database Reference PDB; 1bk6 B; 204; 244; Database Reference PDB; 1ee4 A; 204; 244; Database Reference PDB; 1ee4 B; 204; 244; Database Reference PDB; 1ibr D; 399; 437; Database Reference PDB; 1ibr B; 399; 437; Database Reference PDB; 1qgk A; 399; 437; Database Reference PDB; 1qgr A; 399; 437; Database reference: PFAMB; PB002221; Database reference: PFAMB; PB002617; Database reference: PFAMB; PB004638; Database reference: PFAMB; PB012310; Database reference: PFAMB; PB040528; Database reference: PFAMB; PB041028; Comment: Approx. 40 amino acid repeat. Tandem repeats form super-helix of helices Comment: that is proposed to mediate interaction of beta-catenin with its ligands. Comment: CAUTION: This family does not contain all known armadillo repeats. Number of members: 597
ATP_synt_B_c	PDOC00137	ATP synthase alpha and beta subunits signature	ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), and a catalytic core, called coupling factor CF(1). The former acts as a proton channel; the latter is composed of five subunits, alpha, beta, gamma, delta and epsilon. The sequences of subunits alpha and beta are related and both contain a nucleotide-binding site for ATP and ADP. The beta chain has catalytic activity, while the alpha chain is a regulatory subunit. Vacuolar ATPases [3] (V-ATPases) are responsible for acidifying

Pfam	Prosite	Full Name	Description
			a variety of intracellular compartments in eukaryotic cells. Like F-ATPases, they are oligomeric complexes of a transmembrane and a catalytic sector. The sequence of the largest subunit of the catalytic sector (70 Kd) is related to that of F-ATPase beta subunit, while a 60 Kd subunit, from the same sector, is related to the F-ATPases alpha subunit [4]. Archaebacterial membrane-associated ATPases are composed of three subunits. The alpha chain is related to F-ATPases beta chain and the beta chain is related to F-ATPases alpha chain [4]. A protein highly similar to F-ATPase beta subunits is found [5] in some bacterial apparatus involved in a specialized protein export pathway that proceeds without signal peptide cleavage. This protein is known as flil in Bacillus and Salmonella, Spa47 (mxiB) in Shigella flexneri, HrpB6 in Xanthomonas campestris and yscN in Yersinia virulence plasmids. In order to detect these ATPase subunits, we took a segment of ten amino-acid residues, containing two conserved serines, as a signature pattern. The first serine seems to be important for catalysis - in the ATPase alpha chain at least - as its mutagenesis causes catalytic impairment. Description of pattern(s) and/or profile(s) Consensus pattern P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S [The first S is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL, except for the archaebacterium Sulfolobus acidocaldarius ATPase alpha chain where the first Ser is replaced by Gly. Other sequence(s) detected in SWISS-PROT 37. Note F-ATPase alpha and beta subunits, V-ATPase 70 Kd subunit and the archaebacterial ATPase alpha subunit also contain a copy of the ATP-binding motifs A and B (see <PDOC00017>). Last update November 1997 / Pattern and text revised. References [1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989). [2] Senior A.E. Physiol. Rev. 68:177-231(1988). [3] Nelson N. J. Bioenerg. Biomembr. 21:553-571(1989). [4] Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., Manolson M.F., Poole R.J., Date T., Oshima T., Konishi J., Denda K., Yoshida M. Proc. Natl. Acad. Sci. U.S.A. 86:6661-665(1989). [5] Dreyfus G., Williams A.W., Kawagishi I., MacNab R.M. J. Bacteriol. 175:3131-3138(1993).

Pfam	Prosite	Full Name	Description
ATP-gua Ptrans	PDOC00103	ATP:guanido phosphotransferases active site	ATP:guanido phosphotransferases are a family of structurally and functionally related enzymes [1,2] that reversibly catalyze the transfer of phosphate between ATP and various phosphogens. The enzymes that belongs to this family are: - Creatine kinase (EC 2.7.3.2) (CK) [3,4], which plays an important role in energy metabolism of vertebrates. It catalyzes the reversible transfer of high energy phosphate from ATP to creatine, generating phosphocreatine and ADP. There are at least four different, but very closely related, forms of CK. Two of the CK isozymes are cytosolic: the M (muscle) and B (brain) forms while the two others are mitochondrial. In sea urchin there is a flagellar isozyme, which consists of the triplication of a CK-domain. - Glycocyamine kinase (EC 2.7.3.1) (guanidoacetate kinase), an enzyme that catalyzes the transfer of phosphate from ATP to guanidoacetate. - Arginine kinase (EC 2.7.3.3), an enzyme that catalyzes the transfer of phosphate from ATP to arginine. - Taurocyamine kinase (EC 2.7.3.4), an annelid-specific enzyme that catalyzes the transfer of phosphate from ATP to taurocyamine. - Lombricine kinase (EC 2.7.3.5), an annelid-specific enzyme that catalyzes the transfer of phosphate from ATP to lombricine. - Smc74 [1], a cercaria-specific enzyme from Schistosoma mansoni. This enzyme consists of two CK-related duplicated domains. The substrate(s) specificity of Smc74 is not yet known. A cysteine residue is implicated in the catalytic activity of these enzymes. The region around this active site residue is highly conserved and can be used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern C-P-x(0,1)-[ST]-N-[IL]-G-T [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Stein L.D., Harn D.A., David J.R. J. Biol. Chem. 265:6582-6588(1990). [2] Strong S.J., Ellington W.R. Biochim. Biophys. Acta 1246:197-200(1995). [3] Bessman S.-P., Carpenter C.L. Annu. Rev. Biochem. 54:831-862(1985). [4] Haas R.C., Strauss A.W. J. Biol. Chem. 265:6921-6927(1990).
ATP-synt_D		ATP synthase subunit D	Accession number: PF01813

Pfam	Prosite	Full Name	Description
			Definition: ATP synthase subunit D Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1304 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 157.80 157.80 Noise cutoffs: -79.90 -79.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96324968 Reference Title: Subunit structure and organization of the genes of the A1A0 Reference Title: ATPase from the Archaeon Methanosarcina mazei Go1. Reference Author: Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Reference Location: J Biol Chem 1996;271:18843-18852. Reference Number: [2] Reference Medline: 95132627 Reference Title: A bovine cDNA and a yeast gene (VMA8) encoding the subunit Reference Title: D of the vacuolar H(+)-ATPase. Reference Author: Nelson H, Mandiyan S, Nelson N; Reference Location: Proc Natl Acad Sci U S A 1995;92:497-501. Database Reference INTERPRO; IPR002699; Comment: This is a family of subunit D form various ATP synthases Comment: including V-type H+ transporting and Na+ dependent. Comment: Subunit D is suggested to be an integral part of the Comment: catalytic sector of the V-ATPase [2]. Number of members: 21
ATZ_TRZ		Chlorohydrolase	Accession number: PF01685 Definition: Chlorohydrolase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1192 (release 4.1) Gathering cutoffs: -84 -84 Trusted cutoffs: -74.80 -74.80 Noise cutoffs: -94.30 -94.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96326334 Reference Title: Atrazine chlorohydrolase from Pseudomonas sp strain ADP: Reference Title: gene sequence, enzyme purification, and protein Reference Title: characterization [published erratum appears in J Bacteriol Reference Title: 1999 Jan;181(2):695] Reference Author: de Souza ML, Sadowsky MJ, Wackett LP; Reference Location: J Bacteriol 1996;178:4894-4900. Reference Number: [2] Reference Medline: 96011356 Reference Title: Cloning and expression of the s-triazine hydrolase gene Reference Title: (trzA) from Rhodococcus corallinus and development of Reference Title: Rhodococcus recombinant strains capable of dealkylating and Reference Title: dechlorinating the herbicide atrazine. Reference Author: Shao ZQ, Seffens W, Mulbry W, Behki RM; Reference Location: J Bacteriol 1995;177:5748-5755. Database Reference INTERPRO; IPR002604; Database reference: PFAMB; PB034853; Database reference: PFAMB; PB040603; Comment: This family consist of chlorohydrolase from the ATZ/TRZ family;

Pfam	Prosite	Full Name	Description
			Comment: these enzymes catalyse hydrolytic dechlorination of their substrates. Comment: Atrazine chlorohydrolase (AtzA) from Pseudomonas sp. Swiss:P72156 Comment: catalyses the dechlorination of atrazine to hydroxyatrazine [1]. Comment: s-Triazine hydrolase (TrzA) form R. corallinus Swiss:P72156 Comment: catalyses the deamination and dechlorination of melamine and Comment: deethylsimazine to ammeline and N-ethylammeline [1]. Number of members: 29
B56		Protein phosphatase 2A regulatory B subunit (B56 family)	Accession number: PF01603 Definition: Protein phosphatase 2A regulatory B subunit (B56 family) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_984 (release 4.1) Gathering cutoffs: 11 11 Trusted cutoffs: 17.80 17.80 Noise cutoffs: 5.50 5.50 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96064678 Reference Title: Identification of a new family of protein phosphatase 2A Reference Title: regulatory subunits. Reference Author: McCright B, Virshup DM; Reference Location: J Biol Chem 1995;270:26123-26128. Database Reference INTERPRO; IPR002554; Comment: Protein phosphatase 2A (PP2A) is a major intracellular protein Comment: phosphatase that regulates multiple aspects of cell growth and metabolism. Comment: The ability of this widely distributed heterotrimeric enzyme to act on a Comment: diverse array of substrates is largely controlled by the nature of its Comment: regulatory B subunit. There are multiple families of B subunits (See also Comment: PR55), this family is called the B56 family [1]. Number of members: 34
Bac_export_1		Bacterial export proteins, family 1	Accession number: PF01311 Definition: Bacterial export proteins, family 1 Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1442 (release 3.0) Gathering cutoffs: 25 25 Trusted cutoffs: 37.20 37.20 Noise cutoffs: -95.00 -95.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95113771 Reference Title: Caulobacter FliQ and FliR membrane proteins, required for Reference Title: flagellar biogenesis and cell division, belong to a family Reference Title: of virulence factor export proteins. Reference Author: Zhuang WY, Shapiro L; Reference Location: J Bacteriol 1995;177:343-356. Database Reference INTERPRO; IPR002010; Comment: This family includes the following members; Comment: FliR, MopE, SsaT, YopT, Hrp, HrcT and SpaR Comment: All of these members export proteins, that do not possess signal Comment: peptides, through the membrane. Although the proteins that these

Pfam	Prosite	Full Name	Description
			Comment: exporters move may be different, the exporters are thought to Comment: function in similar ways [1]. Number of members: 29
Band_41	PDOC00566	Band 4.1 family domain signatures and profile	A number of cytoskeletal-associated proteins that associate with various proteins at the interface between the plasma membrane and the cytoskeleton contain a conserved N-terminal domain of about 150 amino-acid residues [1,2, 3]. The proteins in which such a domain is known to exist are listed below. - Band 4.1, which links the spectrin-actin cytoskeleton of erythrocytes to the plasma membrane. Band 4.1 binds with a high affinity to glycophorin and with lower affinity to band 3 protein. - Ezrin (cytovillin or p81), a component of the undercoat of the microvilli plasma membrane. - Moesin, which is probably involved in binding major cytoskeletal structures to the plasma membrane. - Radixin, which seems to play a crucial role in the binding of the barbed end of actin filaments to the plasma membrane in the undercoat of the cell- to-cell adherens junction (AJ). - Talin, which binds with high affinity to vinculin and with low affinity to integrins. Talin is a high molecular weight (270 Kd) cytoskeletal protein concentrated in regions of cell-substratum contact and, in lymphocytes, of cell-cell contacts. - Filopodin, a slime mold protein that binds actin ans which is involved in the control of cell motility and chemotaxis. - Merlin (or schwannomin). Defects in this protein are the cause of type 2 neurofibromatosis (NF2), a predisposition to tumors of the nervous system. - Protein NBL4. - Protein-tyrosine phosphatases PTPN3 (PTP-H1) and PTPN4 (PTP-MEG1). Structurally these two very similar enzymes are composed of a N-terminal band 4.1-like domain followed by a central segment of unknown function and a C-terminal catalytic domain (see <PDOC00323>). They could act at junctions between the membrane and the cytoskeleton. - Protein-tyrosine phosphatases PTPN14 (PEZ or PTP36) and PTP-D1, PTP-RL10 and PTP2E. These phosphatases also consist of a N-terminal band 4.1-like domain and a C-terminal catalytic domain. The central domain seems to contain a SH3-binding domain. - Caenorhabditis elegans protein phosphatase ptp-1. Ezrin, moesin, and radixin are highly related proteins, but the other proteins in which this domain is found do not share any region of similarity outside of the domain. In band 4.1 this domain is known to be important for the interaction with glycophorin, an integral membrane protein. We have developed two signature patterns for this domain, one is based on the

Pfam	Prosite	Full Name	Description
			conserved positions found at the N-terminal extremity of the domain, the second is located in the C-terminal section. Description of pattern(s) and/or profile(s) Consensus pattern W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-[LIVMF]- x(6,8)-[LIVMF]-x(3,5)-F-[FY]-x(2)-[DENS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]-[LIVM]-x(2)-[ACV]- x(2)-[LM]-x(2)-[FY]-G-x-[DENQST]-[LIVMFYS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT 7. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Rees J. jrees@vax.oxford.ac.uk Last update November 1997 / Patterns and text revised; profile added. References [1] Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O. Nature 347:685-689(1990). [2] Funayama N., Nagafuchi A., Sato N., Tsukita S., Tsukita S. J. Cell Biol. 115:1039-1048(1991). [3] Takeuchi K., Kawashima A., Nagafuchi A., Tsukita S. J. Cell Sci. 107:1921-1928(1994).
biotin_lipoyl	PDOC00167; PDOC00168	Biotin-requiring enzymes; 2-oxo acid dehydrogenases acyltransferase component lipoyl binding	Biotin, which plays a catalytic role in some carboxyl transfer reactions, is covalently attached, via an amide bond, to a lysine residue in enzymes requiring this coenzyme [1,2,3,4]. Such enzymes are: - Pyruvate carboxylase (EC 6.4.1.1). - Acetyl-CoA carboxylase (EC 6.4.1.2). - Propionyl-CoA carboxylase (EC 6.4.1.3) - Methylcrotonoyl-CoA carboxylase (EC 6 4.1.4). - Geranoyl-CoA carboxylase (EC 6.4.1.5). - Urea carboxylase (EC 6.3.4.6). - Oxaloacetate decarboxylase (EC 4.1.1.3). - Methylmalonyl-CoA decarboxylase (EC 4.1.1.41). - Glutaconyl-CoA decarboxylase (EC 4.1.1.70). - Methylmalonyl-CoA carboxyl-transferase (EC 2.1.3.1) (transcarboxylase). Sequence data reveal that the region around the biocytin (biotin-lysine) residue is well conserved and can be used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]- x(3)-[LIVM]-x-[SAV] [K is the biotin attachment site]

Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note the domain around the biotin-binding lysine residue is evolutionary related to that around the lipoyl-binding lysine residue of 2-oxo acid dehydrogenase acyltransferases (see <PDOC00168>). Last update November 1997 / Pattern and text revised. References [1] Knowles J.R. Annu. Rev. Biochem. 58:195-221(1989). [2] Samols D., Thronton C.G., Murtif V.L., Kumar G.K., Haase F.C., Wood H.G. J. Biol. Chem. 263:6461-6464(1988). [3] Goss N.H., Wood H.G. Meth. Enzymol. 107:261-278(1984). [4] Shenoy B.C., Xie Y., Park V.L., Kumar G.K., Beegen H., Wood H.G., Samols D. J. Biol. Chem. 267:18407-18412(1992). The 2-oxo acid dehydrogenase multienzyme complexes [1,2] from bacterial and eukaryotic sources catalyze the oxidative decarboxylation of 2-oxo acids to the corresponding acyl-CoA. The three members of this family of multienzyme complexes are: - Pyruvate dehydrogenase complex (PDC). - 2-oxoglutarate dehydrogenase complex (OGDC). - Branched-chain 2-oxo acid dehydrogenase complex (BCOADC). These three complexes share a common architecture: they are composed of multiple copies of three component enzymes - E1, E2 and E3. E1 is a thiamine pyrophosphate-dependent 2-oxo acid dehydrogenase, E2 a dihydrolipamide acyltransferase, and E3 an FAD-containing dihydrolipamide dehydrogenase. E2 acyltransferases have an essential cofactor, lipoic acid, which is covalently bound via a amide linkage to a lysine group. The E2 components of OGCD and BCOACD bind a single lipoyl group, while those of PDC bind either one (in yeast and in Bacillus), two (in mammals), or three (in Azotobacter and in Escherichia coli) lipoyl groups [3]. In addition to the E2 components of the three enzymatic complexes described above, a lipoic acid cofactor is also found in the following proteins: - H-protein of the glycine cleavage system (GCS) [4]. GCS is a multienzyme complex of four protein components, which catalyzes the degradation of glycine. H protein shuttles the methylamine group of glycine from the P protein to the T protein. H-protein from either prokaryotes or eukaryotes binds a single lipoic group.

Pfam	Prosite	Full Name	Description
			- Mammalian and yeast pyruvate dehydrogenase complexes differ from that of other sources, in that they contain, in small amounts, a protein of unknown function - designated protein X or component X. Its sequence is closely related to that of E2 subunits and seems to bind a lipoic group [5]. - Fast migrating protein (FMP) (gene acoC) from Alcaligenes eutrophus [6]. This protein is most probably a dihydrolipamide acyltransferase involved in acetoin metabolism. We developed a signature pattern which allows the detection of the lipoyl- binding site. Description of pattern(s) and/or profile(s) Consensus pattern [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]- [STAVQDN]-x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY] [K is the lipoyl-binding site] Sequences known to belong to this class detected by the pattern ALL Other sequence(s) detected in SWISS-PROT 2. Note the domain around the lipoyl-binding lysine residue is evolutionary related to that around the biotin-binding lysine residue of biotin requiring enzymes (see <PDOC00167>). Last update November 1995 / Text revised. References [1] Yeaman S.J. Biochem. J. 257:625-632(1989). [2] Yeaman S.J. Trends Biochem. Sci. 11:293-296(1986). [3] Russel G.C., Guest J.R. Biochim. Biophys. Acta 1076:225-232(1991). [4] Fujiwara K., Okamura-Ikeda K., Motokawa Y. J. Biol. Chem. 261:8836-8841(1986). [5] Behal R.H , Browning K.S., Hall T.B., Reed L.J. Proc. Natl. Acad. Sci. U.S.A. 86:8732-8736(1989). [6] Priefert H., Hein S., Krueger N., Zeh K., Schmidt B., Steinbuechel A. J. Bacteriol. 173:4056-4071(1991).
Biotin_synth		Biotin synthase	Accession number: PF01792 Definition: Biotin synthase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1407 (release 4.2) Gathering cutoffs: -180 -180 Trusted cutoffs: -176.30 -176.30 Noise cutoffs: -183.90 -183.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96312354 Reference Title: Cloning, sequencing, and characterization of the Bacillus

Pfam	Prosite	Full Name	Description
			Reference Title: subtilis biotin biosynthetic operon Reference Author: Bower S, Perkins JB, Yocum RR, Howitt CL, Rahaim P, Pero J; Reference Location: J Bacteriol 1996;178:4122-4130. Reference Number: [2] Reference Medline: 97074643 Reference Title: Two new members of the bio B superfamily: cloning, Reference Title: sequencing and expression of bio B genes of Methylobacillus Reference Title: flagellatum and Corynebacterium glutamicum. Reference Author: Serebriiskii IG, Vassin VM, Tsygankov YD; Reference Location: Gene 1996;175:15-22. Database Reference INTERPRO; IPR002684; Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 works with flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to convert dethiobiotin to biotin [1]. Comment: Biotin (vitamin H) is a prosthetic group in enzymes catalysing Comment: carboxylation and transcarboxylation reactions [2]. Number of members: 29
BolA		BolA-like protein	Accession number: PF01722 Definition: BolA-like protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1996 (release 4.1) Gathering cutoffs: 23 23 Trusted cutoffs: 23.70 23.70 Noise cutoffs: -16.00 -16.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99291046 Reference Title: The stationary-phase morphogene bolA from Escherichia coli Reference Title: is induced by stress during early stages of growth. Reference Author: Santos JM, Freire P, Vicente M, Arraiano CM; Reference Location: Mol Microbiol 1999;32:789-798. Reference Number: [2] Reference Medline: 90059998 Reference Title: Induction of a growth-phase-dependent promoter triggers Reference Title: transcription of bolA, an Escherichia coli morphogene. Reference Author: Aldea M, Garrido T, Hernandez-Chico C, Vicente M, Kushner Reference Author: SR; Reference Location: EMBO J 1989;8:3923-3931. Database Reference INTERPRO; IPR002634; Comment: This family consist of the morpho-protein BolA from Comment: E. coli and its various homologs. In E. coli over expression of Comment: this protein causes round morphology and may be involved in Comment: switching the cell between elongation and septation systems during Comment: cell division [1]. The expression of BolA is growth rate regulated Comment: and is induced during the transition into the the stationary Comment: phase [1]. BolA is also induced by stress during early stages of Comment: growth [1] and may have a general role in

Pfam	Prosite	Full Name	Description
			stress response. Comment: It has also been suggested that BolA can induce the transcription Comment: of penicillin binding proteins 6 and 5 [2,1]. Number of members: 18
casein_kappa			Accession number: PF00997 Definition: Kappa casein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1298 (release 3.0) Gathering cutoffs: -32 -32 Trusted cutoffs: 16.40 16.40 Noise cutoffs: -73.00 -73.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98072500 Reference Title: Nucleotide sequence evolution at the kappa-casein locus: Reference Title: evidence for positive selection within the family Bovidae. Reference Author: Ward TJ, Honeycutt RL, Derr JN; Reference Location: Genetics 1997;147:1863-1872. Database Reference INTERPRO; IPR000117; Comment: Kappa-casein is a mammalian milk protein involved in a Comment: number of important physiological processes. In the gut, Comment: the ingested protein is split into an insoluble peptide Comment: (para kappa-casein) and a soluble hydrophilic glycopeptide Comment: (caseinomacropeptide). Caseinomacropeptide is responsible Comment: for increased efficiency of digestion, prevention of neonate Comment: hypersensitivity to ingested proteins, and inhibition of Comment: gastric pathogens. Number of members: 56
CAT	PDOC00093	Chloramphenicol acetyltransferase	Chloramphenicol acetyltransferase (CAT) (EC 2.3.1.28) [1] catalyzes the acetyl-CoA dependent acetylation of chloramphenicol (Cm), an antibiotic which inhibits prokaryotic peptidyltransferase activity. Acetylation of Cm by CAT inactivates the antibiotic. A histidine residue, located in the C-terminal section of the enzyme, plays a central role in its catalytic mechanism. We derived a signature pattern from the region surrounding this active site residue. Description of pattern(s) and/or profile(s) Consensus pattern Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H [The second H is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note there is a second family of CAT [2], evolutionary unrelated to the main family described above. These CAT belong to the bacterial hexapeptide-repeat containing-transferases family (see <PDOC00094>). Last update November 1997 / Text revised. References [1]

Pfam	Prosite	Full Name	Description
			Shaw W.V , Leslie A.G.W. Annu. Rev. Biophys. Chem. 20:363-386(1991). [2] Parent R., Roy P.H. J. Bacteriol. 174:2891-2897(1992).
Cation_efflux		Cation efflux family	Accession number: PF01545 Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_232 (release 4.0) Gathering cutoffs: -6 -6 Trusted cutoffs: 6.90 6.90 Noise cutoffs: -19.30 -19.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98361887 Reference Title: Molecular characterization of a chromosomal determinant Reference Title: conferring resistance to zinc and cobalt ions in Reference Title: Staphylococcus aureus. Reference Author: Xiong A, Jayaswal RK; Reference Location: J Bacteriol 1998;180:4024-4029. Reference Number: [2] Reference Medline: 96219090 Reference Title: Cloning and sequence analysis of czc genes in Alcaligenes Reference Title: sp. strain CT14. Reference Author: Kunito T, Kusano T, Oyaizu H, Senoo K, Kanazawa S, Reference Author: Matsumoto S; Reference Location: Biosci Biotechnol Biochem 1996;60:699-704. Database Reference INTERPRO; IPR002524; Database reference: PFAMB; PB038216; Comment: Members of this family are integral membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59
CBD_6		Cellulose binding domain	Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference PDB; 1ulo ; 1; 149; Database Reference PDB; 1ulp ; 1; 149;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1cx1 A; 2; 6; Database Reference PDB; 1ulo ; 150; 152; Database Reference PDB; 1ulp ; 150; 152; Database Reference PDB; 1cx1 A; 7; 151; Database reference: PFAMB; PB012497; Database reference: PFAMB; PB041237; Database reference: PFAMB; PB041605; Number of members: 76
CBFD_NFYB_HMF	PDOC00578	CBF/NF-Y subunits signatures	Diverse DNA binding proteins are known to bind the CCAAT box, a common cis- acting element found in the promoter and enhancer regions of a large number of genes in eukaryotes. Amongst these proteins is one known as the CCAAT-binding factor (CBF) or NF-Y [1]. CBF is a heteromeric transcription factor that consists of two different components both needed for DNA-binding. The HAP protein complex of yeast binds to the upstream activation site of cytochrome C iso-1 gene (CYC1) as well as other genes involved in mitochondrial electron transport and activates their expression. It also recognizes the sequence CCAAT and is structurally and evolutionary related to CBF. The first subunit of CBF, known as CBF-A or NF-YB in vertebrates, HAP3 in budding yeast and as php3 in fission yeast, is a protein of 116 to 210 amino- acid residues which contains a highly conserved central domain of about 90 residues. This domain seems to be involved in DNA-binding; we have developed a signature pattern from its central part. The second subunit of CBF, known as CBF-B or NF-YA in vertebrates, HAP2 in budding yeast and php2 in fission yeast, is a protein of 265 to 350 amino-acid residues which contains a highly conserved region of about 60 residues. This region, called the 'essential core' [2], seems to consist of two subdomains: an N-terminal subunit-association domain and a C-terminal DNA recognition domain. We have developed a signature pattern from a section of the subunit- association domain. Description of pattern(s) and/or profile(s) Consensus pattern C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C Sequences known to belong to this class detected by the pattern ALL CBF-A subunits. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E Sequences known to belong to this class detected by the pattern ALL CBF-B subunits. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I., Benoist C., Mathis D.

Pfam	Prosite	Full Name	Description
			Nucleic Acids Res. 20:1087-1091(1992). [2] Olesen J.T., Fikes J.D., Guarente L. Mol. Cell. Biol. 11:611-619(1991).
CbiX		CbiX	Accession number: PF01903 Definition: CbiX Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -25 -25 Trusted cutoffs: -23.10 -23.10 Noise cutoffs: -35.10 -35.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98416126 Reference Title: Cobalamin (vitamin B12) biosynthesis: identification and Reference Title: characterization of a Bacillus megaterium cobI operon. Reference Author: Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Reference Location: Biochem J 1998;335:159-166. Database Reference INTERPRO; IPR002762; Database reference: PFAMB; PB040604; Database reference: PFAMB; PB040610; Database reference: PFAMB; PB041575; Comment: The function of CbiX is uncertain, however it is found Comment: in cobalamin biosynthesis operons and so may have a Comment: related function. Some CbiX proteins contain a striking Comment: histidine-rich region at their C-terminus, which suggests Comment: that it might be involved in metal chelation [1]. Number of members: 6
cellulase	PDOC00565	Glycosyl hydrolases family 5 signature	The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family A [3] or as the glycosyl hydrolases family 5 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - Endoglucanases from various species and strains of Bacillus. - Butyrivibrio fibrisolvens endoglucanases 1 (end1) and A (celA). - Caldocellum saccharolyticum bifunctional endoglucanase/exoglucanase (celB). This protein consists of two domains; it is the C-terminal domain, which has endoglucanase activity, which belongs to this family. - Clostridium acetobutylicum endoglucanase (eglA). - Clostridium cellulolyticum endoglucanases A (celccA) and D (celccD). - Clostridium cellulovorans endoglucanase B (engB) and D (engD). - Clostridium thermocellum endoglucanases B (celB), C (celC), E (celE), G (celG) and H (celH). - Erwinia chrysanthemi endoglucanase Z (celZ). - Fibrobacter succinogenes endoglucanase 3 (cel-3). - Pseudomonas fluorescens endoglucanase C (celC).

Pfam	Prosite	Full Name	Description
			- Pseudomonas solanacearum endoglucanase (egl). - Robillarda strain Y-20 endoglucanase I. - Ruminococcus albus endoglucanases I (EG-I), A (celA), and B (celB). - Ruminococcus flavefaciens cellodextrinase A (celA). - Ruminococcus flavefaciens endoglucanase E (celE). - Streptomyces lividans endoglucanase. - Thermomonospora fusca endoglucanase E-5 (celE). - Trichoderma reesei endoglucanase II (EGLII). - Xanthomonas campestris endoglucanase (engxcA). As well as: - Baker's yeast glucan 1,3-beta-glucosidase I/II (EC 3.2.1.58) (EXG1). - Baker's yeast glucan 1,3-beta-glucosidase 2 (EC 3.2.1.58) (EXG2). - Baker's yeast sporulation-specific glucan 1,3-beta-glucosidase (SPR1). - Caldocellum saccharolyticum beta-mannanase (EC 3.2.1.78) (manA). - Yeast hypothetical protein YBR056w. - Yeast hypothetical protein YIR007w. One of the conserved regions in these enzymes contains a conserved glutamic acid residue which is potentially involved [5] in the catalytic mechanism. We use this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]- [RHDNSTLIVFY] [E is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Robillarda Y-20 endoglucanase I whose sequence is known to be incorrect and yeast YBR056w. Other sequence(s) detected in SWISS-PROT 22. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1997 / Pattern and text revised. References [1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55.303-315(1991). [3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989). [4] Henrissat B. Biochem. J. 280:309-316(1991). [5] Py B., Bortoli-German I., Haiech J., Chippaux M., Barras F. Protein Eng. 4:325-333(1991). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt
CH	PDOC00019	Actinin-type actin-binding domain signatures	Alpha-actinin is a F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures [1]. The actin-binding

Pfam	Prosite	Full Name	Description
			domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]: - In the beta chain of spectrin (or fodrin). - In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane. - In the slime mold gelation factor (or ABP-120). - In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to membrane glycoproteins. - In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs from the above proteins in that it contains two tandem copies of the actin- binding domain and that these copies are located in the C-terminal part of the protein. We selected two conserved regions as signature patterns for this type of domain. The first of this region is located at the beginning of the domain, while the second one is located in the central section and has been shown to be essential for the binding of actin. Description of pattern(s) and/or profile(s) Consensus pattern [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 25. Consensus pattern [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x- [DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x- [LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988). [2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991). [3] Dubreuil R.R. BioEssays 13:219-226(1991).
chitinase_2	PDOC00839	Chitinases family 18 active site	Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 18 (also known as classes III or V) groups a variety of proteins: a) Chitinases from:

Pfam	Prosite	Full Name	Description
			- Prokaryotes such as Alteromonas, Bacillus, Serratia, Streptomyces, etc. - Plants such as Arabidopsis, cucumber, bean, tobacco, etc. - Fungi such as Aphanocladium, Rhizopus, Saccharomyces, etc. - Nematode (Brugia malayi). - Insects (Manduca sexta). - Baculoviruses (Autographa Californica Nuclear Polyhedrosis virus). b) Other proteins: - Hevamine, a rubber tree protein with chitinase and lysozyme activities. - Kluyveromyces lactis killer toxin alpha subunit, which acts as a chitinase. - Flavobacterium and Streptomyces endo-beta-N-acetylglucosaminidases (EC 3.2. 1.96). - Mammalian di-N-acetylchitobiase which is involved in the degradation of asparagine-linked glycoproteins. - Human cartilage glycoprotein Gp-39. - Jack bean concanavalin B (conB), a protein that has lost its catalytic activity. Site directed mutagenesis experiments [3] and crystallographic data [4,5] have shown that a conserved glutamate is involved in the catalytic mechanism and probably acts as a proton donor. This glutamate is at the extremity of the best conserved region in these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for conB which has a Gln instead of the active site Glu. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Neuhaus J.-M. jean-marc.neuhaus@bota.unine.ch Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1997 / Text revised. References [1] Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992). [2] Henrissat B Biochem. J. 280:309-316(1991). [3] Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H., Uchida M., Tanaka H. J. Biol. Chem. 268:18567-18572(1993). [4] Perrakis A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S., Vorgias C.E. Structure 2:1169-1180(1994). [5] van Scheltinga A.C.T., Kalk K.H., Beintema J.J., Dijkstra B.W. Structure 2:1181-1189(1994).

Pfam	Prosite	Full Name	Description
			[E1] http://www.expasy.ch/cgi-bi/lists?glycosid.txt
Choline_kinase		Choline/ethanolamine kinase	Accession number: PF01633 Definition: Choline/ethanolamine kinase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1165 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 242.90 242.90 Noise cutoffs: -85.90 -85.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98175949 Reference Title: Expression, purification, and characterization of choline Reference Title: kinase, product of the CKI gene from Saccharomyces Reference Title: cerevisiae. Reference Author: Kim KH, Voelker DR, Flocco MT, Carman GM; Reference Location: J Biol Chem 1998;273:6844-6852. Database Reference INTERPRO; IPR002573; Comment: Choline kinase catalyses the committed step in the synthesis of Comment: phosphatidylcholine by the CDP-choline pathway [1]. Number of members: 22
Chorion		Chorion protein	Accession number: PF01723 Definition: Chorion protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1914 (release 4.1) Gathering cutoffs: -46 -46 Trusted cutoffs: -43.70 -43.70 Noise cutoffs: -49.00 -49.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95333194 Reference Title: Sequence analysis of a small early chorion gene subfamily Reference Title: interspersed within the late gene locus in Bombyx mori. Reference Author: Kravariti L, Lecanidou R, Rodakis GC; Reference Location: J Mol Evol 1995;41:24-33. Reference Number: [2] Reference Medline: 86313609 Reference Title: Evolution of the silk moth chorion gene superfamily: gene Reference Title: families CA and CB. Reference Author: Lecanidou R, Rodakis GC, Eickbush TH, Kafatos FC; Reference Location: Proc Natl Acad Sci U S A 1986;83:6514-6518. Database Reference INTERPRO; IPR002635; Database reference: PFAMB; PB009425; Comment: This family consists of the chorion superfamily proteins Comment: classes A, B, CA, CB and high-cysteine HCB from silk, Comment: gypsy and polyphemus moths. Comment: The chorion proteins make up the moths egg shell a complex Comment: extracellular structure [2]. Number of members: 35
Chorismate_mut		Chorismate mutase	Accession number. PF01817 Definition: Chorismate mutase Author: Bateman A Alignment method of seed: Manual

Pfam	Prosite	Full Name	Description
			Source of seed members: PSI-BLAST 1ecm Gathering cutoffs: 5 5 Trusted cutoffs: 5.10 5.10 Noise cutoffs: -19.90 -19.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95062155 Reference Title: The crystal structure of allosteric chorismate mutase at Reference Title: 2.2-A resolution. Reference Author: Xue Y, Lipscomb WN, Graf R, Schnappauf G, Braus G; Reference Location: Proc Natl Acad Sci U S A 1994;91:10814-10818. Reference Number: [2] Reference Medline: 98307941 Reference Title: Tyrosine and tryptophan act through the same binding site Reference Title: at the dimer interface of yeast chorismate mutase. Reference Author: Schnappauf G, Krappmann S, Braus GH; Reference Location: J Biol Chem 1998;273:17012-17017. Reference Number: [3] Reference Medline: 98165805 Reference Title: Chorismate mutase-prephenate dehydratase from Escherichia Reference Title: coli. Study of catalytic and regulatory domains using Reference Title: genetically engineered proteins. Reference Author: Zhang S, Pohnert G, Kongsaeree P, Wilson DB, Clardy J, Reference Author: Ganem B; Reference Location: J Biol Chem 1998;273:6248-6253. Database Reference: SCOP; 1csm; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002701; Database Reference PDB; 1ecm B; 6; 89; Database Reference PDB; 1ecm A; 5; 89; Database Reference PDB; 1csm A; 133; 162; Database Reference PDB; 3csm A, 133; 243; Database Reference PDB; 3csm B; 133; 243; Database Reference PDB; 4csm A; 133; 243; Database Reference PDB; 4csm B; 133; 243; Database Reference PDB; 5csm A; 133; 243; Database Reference PDB; 2csm A; 133; 246; Comment: Chorismate mutase EC:5.4.99.5 catalyses the conversion of Comment: chorismate to prephenate in the pathway of tyrosine and Comment: phenylalanine biosynthesis. This enzyme is negatively Comment: regulated by tyrosine, tryptophan and phenylalanine [2,3]. Number of members: 28
CN_hydrolase	PDOC00712; PDOC00943	Nitrilases / cyanide hydratase signatures, Uncharacterized protein family UPF0012 signature	Nitrilases (EC 3.5.5.1) are enzymes that convert nitriles into their corresponding acids and ammonia. They are widespread in microbes as well as in plants where they convert indole-3-acetonitrile to the hormone indole-3- acetic acid. A conserved cysteine has been shown [1,2] to be essential for enzyme activity; it seems to be involved in a nucleophilic attack on the nitrile carbon atom. Cyanide hydratase (EC 4.2.1.66) converts HCN to formamide. In phytopathogenic fungi, it is used to avoid the toxic effect of cyanide released by wounded plants [3]. The sequence of cyanide hydrolase is evolutionary related to that

Pfam	Prosite	Full Name	Description
			of nitrilases. Yeast hypothetical proteins YIL164c and YIL165c also belong to this family. As signature patterns for these enzymes, we selected two conserved regions. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central section. Description of pattern(s) and/or profile(s) Consensus pattern G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Patterns and text revised. References [1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993). [2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yamada H. J. Biol. Chem. 267:20746-20751(1992). [3] Wang P., Vanetten H.D. Biochem. Biophys. Res. Commun. 187:1048-1054(1992). The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome X hypothetical protein YJL126w. - Yeast chromosome XII hypothetical protein YLR351c. - Fission yeast hypothetical protein SpAC26A3.11. - Escherichia coli hypothetical protein ybeM. - Bacillus subtilis hypothetical protein yhcX. - Mycobacterium tuberculosis hypothetical protein MtCY20G9.06c. - Synechocystis strain PCC 6803 hypothetical protein sll0601. - A Pseudomonas fluorescens hypothetical protein in pqqF 5'region. - A Staphylococcus hypothetical protein in agr operon. Except for yhcX which is larger, these are protein of about 30 to 35 Kd which contain, in their central section, a well conserved region centered on a cysteine residue. Description of pattern(s) and/or profile(s) Consensus pattern [GTA]-x(2)-[IVT]-C-Y-D-[LIVM]-x-F-P-x(9)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / First entry. References

<table>
<tr><th>Pfam</th><th>Prosite</th><th>Full Name</th><th>Description</th></tr>
<tr><td></td><td></td><td></td><td>[1]
Bairoch A.
Unpublished observations (1995).</td></tr>
<tr><td>CorA</td><td></td><td>CorA-like Mg2+ transporter protein</td><td>Accession number: PF01544
Definition: CorA-like Mg2+ transporter protein
Author: Bateman A
Alignment method of seed: Clustalw
Source of seed members: Pfam-B_944 (release 4.0)
Gathering cutoffs: -62 -62
Trusted cutoffs: -5.90 -5.90
Noise cutoffs: -86.20 -86.20
HMM build command line: hmmbuild -F HMM SEED
HMM build command line: hmmcalibrate --seed 0 HMM
Reference Number: [1]
Reference Medline: 98448512
Reference Title: The CorA magnesium transporter gene family.
Reference Author: Kehres DG, Lawyer CH, Maguire ME;
Reference Location: Microb Comp Genomics 1998;3:151-169.
Reference Number: [2]
Reference Medline: 99003207
Reference Title: The CorA Mg2+ transport protein of Salmonella typhimurium.
Reference Title: Mutagenesis of conserved residues in the third membrane
Reference Title: domain identifies a Mg2+ pore.
Reference Author: Smith RL, Szegedy MA, Kucharski LM, Walker C, Wiet RM,
Reference Author: Redpath A, Kaczmarek MT, Maguire ME;
Reference Location: J Biol Chem 1998;273:28663-28669.
Database Reference INTERPRO; IPR002523;
Database reference: PFAMB; PB041399;
Comment: The CorA transport system is the primary Mg2+ influx system of Salmonella
Comment: typhimurium and Escherichia coli. CorA is virtually ubiquitous in the
Comment: Bacteria and Archaea. There are also eukaryotic relatives of this protein
Number of members: 25</td></tr>
<tr><td>Cys_knot</td><td>PDOC00234</td><td>Glycoprotein hormones beta chain signatures</td><td>Glycoprotein hormones [1,2] (or gonadotropins) are a family of proteins which
include the mammalian hormones follitropin (FSH), lutropin (LSH), thyrotropin
(TSH) and chorionic gonadotropin (CG), as well as at least two forms of fish
gonadotropins. All these hormones consist of two glycosylated chains (alpha
and beta). In mammalian gonadotropins, the alpha chain is identical in the
four types of hormones but the beta chains, while homologous, are different.

The beta chains are proteins of about 100 to 140 amino acid residues which
contain twelve conserved cysteines all involved in disulfide bonds [3], as
shown in the following schematic representation.

+---------------------+
| +-----------|-----------------------------+
| +-|-----------|--------+ |
| | | **** | | **************

xxxCxxxxxxxCxCxxCxCxxxxxxxCxxxxxxxxCxxxxxxxCxCxCxxCxx xxxCxxxxxxxxxxx
        || ||| |
        || || +--+
        +-|--------------------+ |
          +-----------------------+

'C': conserved cysteine involved in a disulfide bond.
'*': position of the patterns.</td></tr>
</table>

Pfam	Prosite	Full Name	Description
			We have developed two patterns for these hormones. The first one, located in the N-terminal section, is a region which has been said to be involved in the association between the two chains of the hormones. The second pattern consists of a cluster of five conserved cysteines in the C-terminal section. Description of pattern(s) and/or profile(s) Consensus pattern C-[STAGM]-G-[HFYL]-C-x-[ST] [The two C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern ALL, except for rat beta-FSH which has Glu in position 2 of the pattern. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [PA]-V-A-x(2)-C-x-C-x(2)-C-x(4)-[STD]-[DEY]-C-x(6,8)- [PGSTAVM]-x(2)-C [The five C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Lapthorn A. adrian@chem.gla.ac.uk Last update July 1998 / Patterns and text revised. References [1] Pierce J.G., Parsons T.F. Annu. Rev. Biochem. 50:465-495(1981). [2] Stockell Hartree A., Renwick A.G.C. Biochem. J. 287:665-679(1992). [3] Lapthorn A.J., Harris D.C., Littlejohn A., Lustbader J.W., Canfield R.E., Machin K.J., Morgan F.J., Isaacs N.W. Nature 369:455-461(1994).
cytochrome_b_C	PDOC00171	Cytochrome b/b6 signatures	In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex. Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB and petD genes. The sequence of petB is colinear with the N-terminal part of mitochondrial cytochrome b, while petD corresponds to the C-terminal part. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups.

<table>
<tr><th>Pfam</th><th>Prosite</th><th>Full Name</th><th>Description</th></tr>
<tr><td></td><td></td><td></td><td>Apart from regions around some of the histidine heme ligands, there are a few
conserved regions in the sequence of b/b6. The best conserved of these regions
includes an invariant P-E-W triplet which lies in the loop that separates the
fifth and sixth transmembrane segments. It seems to be important for electron
transfer at the ubiquinone redox site - called Qz or Qo (where o stands for
outside) - located on the outer side of the membrane.

A schematic representation of the structure of cytochrome b/b6 is shown below.

+---Fe-b562----+
| +---Fe-b566--|-+
| | | |

xxxxxxxxxxxHxHxxxxxxxxxxxxxHxHxxxxxxxxxxPEWxxxxxxxxxxxxxx xxxxx
<------------------Cytochrome-b--------------------------->
<----Cytochrome-b6-petB---------><--Cytochrome-b6-petD----->

We developed two signature patterns for cytochrome b/b6. The first includes
the first conserved histidine of b/b6, which is a heme b562 ligand; the second
includes the conserved PEW triplet.

Description of pattern(s) and/or profile(s)

Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H
[H is a heme b562 ligand]
Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences.
Other sequence(s) detected in SWISS-PROT 15.

Consensus pattern P-[DE]-W-[FY]-[LFY](2)
Sequences known to belong to this class detected by the pattern ALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b.
Other sequence(s) detected in SWISS-PROT 1.
Last update
November 1995 / Patterns and text revised.
References
[1]
Howell N.
J. Mol. Evol. 29:157-169(1989).

[2]
Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A.
Biochim. Biophys. Acta 1143:243-271(1993).</td></tr>
<tr><td>cytochrome_b_N</td><td>PDOC00171</td><td>Cytochrome b/b6 signatures</td><td>In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is
a component of respiratory chain complex III (EC 1.10.2.2) - also known as the
bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and
cyanobacteria, there is a analogous protein, cytochrome b6, a component of the
plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f
complex.

Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400
amino acid residues that probably has 8 transmembrane segments. In plants and</td></tr>
</table>

<table>
<tr><th>Pfam</th><th>Prosite</th><th>Full Name</th><th>Description</th></tr>
<tr><td></td><td></td><td></td><td>cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB
and petD genes. The sequence of petB is colinear with the N-terminal part of
mitochondrial cytochrome b, while petD corresponds to the C-terminal part.
Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566.
Four conserved histidine residues are postulated to be the ligands of the
iron atoms of these two heme groups.

Apart from regions around some of the histidine heme ligands, there are a few
conserved regions in the sequence of b/b6. The best conserved of these regions
includes an invariant P-E-W triplet which lies in the loop that separates the
fifth and sixth transmembrane segments. It seems to be important for electron
transfer at the ubiquinone redox site - called Qz or Qo (where o stands for
outside) - located on the outer side of the membrane.

A schematic representation of the structure of cytochrome b/b6 is shown below.

+---Fe-b562----+
| +---Fe-b566--|-+
| | | |

xxxxxxxxxxxHxHxxxxxxxxxxxxHxHxxxxxxxxxxPEWxxxxxxxxxxxxxxxxxx
<-----------------Cytochrome-b--------------------------->
<----Cytochrome-b6-petB---------><--Cytochrome-b6-petD----->

We developed two signature patterns for cytochrome b/b6. The first includes
the first conserved histidine of b/b6, which is a heme b562 ligand; the second
includes the conserved PEW triplet.

Description of pattern(s) and/or profile(s)

Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H
[H is a heme b562 ligand]
Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences.
Other sequence(s) detected in SWISS-PROT 15.

Consensus pattern P-[DE]-W-[FY]-[LFY](2)
Sequences known to belong to this class detected by the pattern ALL, except for Odocoileus hemionus (mule deer) and Paramecium tetraurelia cytochrome b.
Other sequence(s) detected in SWISS-PROT 1.
Last update
November 1995 / Patterns and text revised.
References
[1]
Howell N.
J. Mol. Evol. 29:157-169(1989).

[2]
Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A.
Biochim. Biophys. Acta 1143:243-271(1993).</td></tr>
<tr><td>cytochrome_c</td><td>PDOC00169</td><td>Cytochrome c family heme-binding site signature</td><td>In proteins belonging to cytochrome c family [1], the heme group is covalently
attached by thioether bonds to two conserved cysteine residues. The consensus</td></tr>
</table>

Pfam	Prosite	Full Name	Description
			sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c. Description of pattern(s) and/or profile(s) Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the pattern ALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT 454. Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 ! Last update June 1992 / Text revised. References [1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).
DAHP_synth_2		Class-II DAHP synthetase family	Members of this family are aldolase enzymes that catalyse the first step of the shikimate pathway. These polypeptides can be useful in the synthesis of aromatic compounds, such as amino acids, antibiotics, secondary metabolites, etc. Such synthesis can occur either in vitro or in vivo.
Dala_Dala_ligas		D-ala D-ala ligase	Accession number: PF01820 Definition: D-ala D-ala ligase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST 2dln Gathering cutoffs: 25 25 Trusted cutoffs: 44.90 26.60 Noise cutoffs: 21.50 18.90 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97207065 Reference Title: D-alanine:D-alanine ligase: phosphonate and phosphinate Reference Title: intermediates with wild type and the Y216F mutant. Reference Author: Fan C, Park IS, Walsh CT, Knox JR; Reference Location: Biochemistry 1997;36.2531-2538. Database Reference: SCOP; 2dln; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000291; Database Reference PDB; 1iov ; 3; 303; Database Reference PDB; 1iow ; 3; 303; Database Reference PDB; 2dln ; 3; 303; Comment: This family contains D-alanine--D-alanine ligase enzymes EC:6.3.2.4. Number of members: 80
DHPS	PDOC00630	Dihydropteroate synthase signatures	All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes that are involved in the biosynthesis of folates are therefore the target of a variety of

Pfam	Prosite	Full Name	Description
			antimicrobial agents such as trimethoprim or sulfonamides. Dihydropteroate synthase (EC 2.5.1.15) (DHPS) catalyzes the condensation of 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate to para-aminobenzoic acid to form 7,8-dihydropteroate. This is the second step in the three steps pathway leading from 6-hydroxymethyl-7,8-dihydropterin to 7,8-dihydrofolate DHPS is the target of sulfonamides which are substrates analog that compete with para-aminobenzoic acid. Bacterial DHPS (gene sul or folP) [1] is a protein of about 275 to 315 amino acid residues which is either chromosomally encoded or found on various antibiotic resistance plasmids. In the lower eukaryote Pneumocystis carinii, DHPS is the C-terminal domain of a multifunctional folate synthesis enzyme (gene fas) [2]. We developed two signature patterns for DHPS, the first signature is located in the N-terminal section of these enzymes, while the second signature is located in the central section. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Slock J., Stahly D.P., Han C.-Y., Six E.W., Crawford I.P J. Bacteriol. 172:7211-7226(1990). [2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).
DHquinase_I	PDOC00788	Dehydroquinase class I active site	3-dehydroquinate dehydratase (EC 4.2.1.10), or dehydroquinase, catalyzes the conversion of 3-dehydroquinate into 3-dehydroshikimate. It is the third step in the shikimate pathway for the biosynthesis of aromatic amino acids from chorismate. Two classes of dehydroquinases exist, known as types I and II. The best studied type I enzyme is from Escherichia coli (gene aroD) and related bacteria where it is a homodimeric protein of a chain of about 250 residues. In fungi, dehydroquinase is part of a multifunctional enzyme which catalyzes five consecutive steps in the shikimate pathway. In aroD, it has been shown

Pfam	Prosite	Full Name	Description
			[1] that a histidine is involved in the catalytic mechanism; we used the region around this residue as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern D-[LIVM]-[DE]-[LIVMN]-x(18,20)-[LIVM](2)-x-[SC]-[NHY]-H- [DN] [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Deka R.K., Kleanthous C., Coggins J.R. J. Biol. Chem. 267:22237-22242(1992).
Diphthamide_syn		Putative diphthamide synthesis protein	Accession number: PF01866 Definition: Putative diphthamide synthesis protein Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 44.90 44.90 Noise cutoffs: -174.70 -174.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96183112 Reference Title: A cDNA from the ovarian cancer critical region of deletion Reference Title: on chromosome 17p13.3. Reference Author: Phillips NJ, Zeigler MR, Deaven LL; Reference Location: Cancer Lett 1996;102:85-90. Reference Number: [2] Reference Medline: 94010339 Reference Title: Diphthamide synthesis in Saccharomyces cerevisiae: Reference Title: structure of the DPH2 gene. Reference Author: Mattheakis LC, Sor F, Collier RJ; Reference Location: Gene 1993;132:149-154. Database Reference INTERPRO; IPR002728; Comment: Swiss:Q16439 is a candidate tumour suppressor gene [1]. DPH2 from Comment: yeast Swiss:P32461 [2], which confers resistance to diphtheria toxin Comment: has been found to be involved in diphthamide synthesis. Diphtheria Comment: toxin inhibits eukaryotic protein synthesis by ADP-ribosylating Comment: diphthamide, a posttranslationally modified histidine residue present Comment: in EF2. The exact function of the members of this family is Comment: unknown. Number of members: 12
Disintegrin	PDOC00351	Disintegrins signature	Disintegrins [1,2] are snake venom proteins which inhibit fibrinogen interaction with platelet receptors expressed on the glycoprotein IIb-IIIa complex. They act by binding to the integrin glycoprotein IIb-IIIa receptor on the platelet surface and inhibit aggregation induced by ADP, thrombin, platelet-activating factor and collagen. Disintegrins are peptides of about 70 amino acid residues that contain many cysteines all involved in disulfide bonds [3]. Disintegrins contain an Arg-

Pfam	Prosite	Full Name	Description

Gly-Asp (RGD) sequence, a recognition site of many adhesion proteins. The RGD
sequence of disintegrins is postulated to interact with the glycoprotein IIb-
IIIa complex.

The sequences of disintegrins from different snake species are known. These
proteins are known as: albolabrin, applagin, barbourin, batroxostatin,
bitistatin, echistatin, elegantin, eristicophin, flavoridin, halysin, kistrin,
tergeminin and triflavin.

Some other proteins are known to contain a disintegrin domain:

- Some snake venom zinc metalloproteinases [4] consist of an N-terminal
  catalytic domain fused to a disintegrin domain. Such is the case for
  trimerelysin I (HR1B), atrolysin e (Ht-e) and trigramin. It has been
  suggested that these proteinases are able to cleave themselves from the
  disintegrin domains and that the latter may arise from such a post-
  translational processing.
- The beta-subunit of guinea pig sperm surface protein PH30 [5]. PH30 is a
  protein involved in sperm-egg fusion. The beta subunit contains a
  disintegrin at the N-terminal extremity.
- Mammalian epididymial protein 1 (EAP I) [6]. EAP I is associated with the
  sperm membrane and may play a role in sperm maturation. Structurally, EAP I
  consists of an N-terminal domain, followed by a zinc metalloproteinase
  domain, a disintegrin domain, and a large C-terminal domain that contains a
  transmembrane region.

The schematic representation of the structure of a typical disintegrin is
shown below:

```
                     +---+
  +--------+        +----|---|-------------------+
  |      |          |  |  |                |
xxxxxCxCxxxxxxCCxxxxCxxxxxxxCxxxxCCxxCxxxxxxxxCxxxRGD
xxxxxCxxxxxxCxxxxxxx
    |    |  |      ******|***********|*          |
    +-------+  +-------------+      +-----------------+
```

'C': conserved cysteine involved in a disulfide bond.
'*': position of the pattern.

As a signature pattern for disintegrins, we selected a conserved central
region that contains five of the cysteines involved in disulfide bonds.

Description of pattern(s) and/or profile(s)

Consensus pattern C-x(2)-G-x-C-C-x-[NQRS]-C-x-[FM]-x(6)-C-[RK]
Sequences known to belong to this class detected by the pattern ALL.
Other sequence(s) detected in SWISS-PROT NONE.
Last update
December 1992 / Pattern and text revised.

Pfam	Prosite	Full Name	Description
			References [1] Williams J., Rucinski B., Holt J., Niewiarowski S. Biochim. Biophys. Acta 1039:81-89(1990). [2] Dennis M.S., Henzel W.J., Pitti R.M., Lipari M.T., Napier M.A., Deisher T.A., Bunting S., Lazarus R.A. Proc. Natl. Acad. Sci. U.S.A. 87:2471-2475(1990). [3] Calvete J.J., Schaefer W., Soszka T., Lu W., Cook J.J., Jameson B.A., Niewiarowski S. Biochemistry 30:5225-5229(1991). [4] Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992). [5] Blobel C.P., Wolfsberg T.G., Turck C.W., Myles D.G., Primakoff P., White J.M. Nature 356:248-252(1992). [6] Perry A.C.F., Jones R., Barker P.J., Hall L. Biochem. J. 286:671-675(1992).
DLH		Dienelactone hydrolase family	Accession number: PF01738 Definition: Dienelactone hydrolase family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_757 (release 4.2) Gathering cutoffs: 15 0 Trusted cutoffs: 15.60 3.10 Noise cutoffs: 14.40 14.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 90339491 Reference Title: Refined structure of dienelactone hydrolase at 1.8 A. Reference Author: Pathak D, Ollis D; Reference Location: J Mol Biol 1990;214:497-525. Database Reference: SCOP; 1din; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002925; Database Reference PDB; 1din ; 16; 232; Database reference: PFAMB; PB004640; Database reference: PFAMB; PB041131; Database reference: PFAMB; PB041469; Number of members: 42
DNA_mis_repair	PDOC00057	DNA mismatch repair proteins mutL / hexB / PMS1 signature	Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related. These proteins are: - Escherichia coli and Salmonella typhimurium mutL protein [2]. MutL is required for dam-dependent methyl-directed DNA repair. - Streptococcus pneumoniae hexB protein [3]. The Hex system is nick directed. - Yeast proteins PMS1 and MLH1 [4]. - Human protein MLH1 [5] which is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).

Pfam	Prosite	Full Name	Description
			As a signature pattern for this class of mismatch repair proteins we selected a perfectly conserved heptapeptide which is located in the N-terminal section of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern G-F-R-G-E-A-L Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987). [2] Mankovich J.A., McIntyre C.A., Walker G.C. J. Bacteriol. 171:5325-5331(1989). [3] Prudhomme M., Martin B., Mejean V., Claverys J.-P. J. Bacteriol. 171:5332-5338(1989). [4] Prolla T.A., Christie D., Liskay R.M. Mol. Cell. Biol. 14:407-415(1994). [5] Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G., Lescoe M.K., Kane M., Earibino C., Lipford J , Linblom A., Tannergard P., Bollag R.J., Godwin A.R., Ward D.C., Nordenskjold M., Fishel R., Kolodner R.D., Liskay R.M. Nature 368:258-261(1994).
DNA_primase_S		DNA primase small subunit	Accession number: PF01896 Definition: DNA primase small subunit Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 198.40 198.40 Noise cutoffs: -120.80 -120.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91219475 Reference Title: Mutations in conserved yeast DNA primase domains impair DNA Reference Title: replication in vivo. Reference Author: Francesconi S, Longhese MP, Piseri A, Santocanale C, Reference Author: Lucchini G, Plevani P; Reference Location: Proc Natl Acad Sci U S A 1991;88:3877-3881. Database Reference INTERPRO; IPR002755; Comment: DNA primase synthesizes the RNA primers for the Okazaki Comment: fragments in lagging strand DNA synthesis. DNA primase Comment: is a heterodimer of large and small subunits. Number of members: 14
DnaB		DnaB-like helicase	Members of this family are comprise DNA replication enzymes which unwind the helix. Generally, such polypeptide are ATPases which move at the replication fork, disrupting hydrogen bonds. Such proteins are use for DNA replication in vivo and/or in vitro.
DnaJ_C		DnaJ C terminal region	Accession number: PF01556

Pfam	Prosite	Full Name	Description
			Definition: DnaJ C terminal region Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_342 (release 4.0) Gathering cutoffs: -24 -24 Trusted cutoffs: -22.60 -22.60 Noise cutoffs: -25.50 -25.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98308847 Reference Title: The J-domain family and the recruitment of chaperone power. Reference Author: Kelley WL; Reference Location: Trends Biochem Sci 1998;23:222-227. Database Reference INTERPRO; IPR002939; Database reference: PFAMB; PB013976; Comment: This family consists of the C terminal region form the DnaJ Comment: protein. Although the function of this region is unknown, it Comment: is always found associated with DnaJ and DnaJ_CXXCXGXG. Comment: DnaJ is a chaperone associated with the Hsp70 heat-shock Comment: system involved in protein folding and renaturation after stress. Number of members: 116
DnaJ_CXXCXGXG	PDOC00553	dnaJ domains signatures and profile	The prokaryotic heat shock protein dnaJ interacts with the chaperone hsp70- like dnaK protein [1]. Structurally, the dnaJ protein consists of an N- terminal conserved domain (called 'J' domain) of about 70 amino acids, a glycine-rich region ('G' domain') of about 30 residues, a central domain containing four repeats of a CXXCXGXG motif ('CRR' domain) and a C-terminal region of 120 to 170 residues. Such a structure is shown in the following schematic representation: +------------+-+-------+-----+-----------+--------------------------------+ | N-terminal | | Gly-R | | CXXCXGXG | C-terminal | +-----------+-+-------+-----+-----------+--------------------------------+ It has been shown [2] that the 'J' domain as well as the 'CRR' domain are also found in other prokaryotic and eukaryotic proteins which are listed below. a) Proteins containing both a 'J' and a 'CRR' domain: - Yeast protein MAS5/YDJ1 which seems to be involved in mitochondrial protein import. - Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding. - Yeast protein SCJ1, involved in protein sorting. - Yeast protein XDJ1. - Plants dnaJ homologs (from leek and cucumber). - Human HDJ2, a dnaJ homolog of unknown function. - Yeast hypothetical protein YNL077w. b) Proteins containing a 'J' domain without a 'CRR' domain: - Rhizobium fredii nolC, a protein involved in cultivar-specific nodulation of soybean. - Escherichia coli cbpA [3], a protein that binds curved DNA. - Yeast protein SEC63/NPL1, important for protein assembly

Pfam	Prosite	Full Name	Description
			into the endoplasmic reticulum and the nucleus. - Yeast protein SIS1, required for nuclear migration during mitosis. - Yeast protein CAJ1. - Yeast hypothetical protein YFR041c. - Yeast hypothetical protein YIR004w. - Yeast hypothetical protein YJL162c. - Plasmodium falciparum ring-infected erythrocyte surface antigen (RESA). RESA, whose function is not known, is associated with the membrane skeleton of newly invaded erythrocytes. - Human HDJ1. - Human HSJ1, a neuronal protein. - Drosophila cysteine-string protein (csp). We developed a signature pattern for the 'J' domain, based on conserved positions in the C-terminal half of this domain. We also developed a pattern for the 'CRR' domain, based on the first two copies of that motif. We also developed a profile for the 'J' domain. Description of pattern(s) and/or profile(s) Consensus pattern [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)- [KR]-x(2)-[FYI] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C- x(2,3)-C-x-G-x-G Sequences known to belong to this class detected by the pattern ALL, except for yeast XDJ1. Other sequence(s) detected in SWISS-PROT 8. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Kelley W. kelley@medecine.unige.ch Last update July 1998 / Patterns and text revised. References [1] Cyr D.M., Langer T., Douglas M.G. Trends Biochem. Sci. 19:176-181(1994). [2] Bork P., Sander C., Valencia A., Bukau B. Trends Biochem. Sci. 17:129-129(1992). [3] Ueguchi C., Kaneda M., Yamada H., Mizuno T. Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).
dNK		Deoxynucleoside kinase	Accession number: PF01712 Definition: Deoxynucleoside kinase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1744 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 47.50 47.50

Pfam	Prosite	Full Name	Description
			Noise cutoffs: -5.40 -5.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97236800 Reference Title: Cloning of the cDNA and chromosome localization of the gene Reference Title: for human thymidine kinase 2. Reference Author: Johansson M, Karlsson A; Reference Location: J Biol Chem 1997;272:8454-8458. Reference Number: [2] Reference Medline: 96293511 Reference Title: Cloning and expression of human deoxyguanosine kinase cDNA. Reference Author: Johansson M, Karlsson A; Reference Location: Proc Natl Acad Sci U S A 1996;93:7258-7262. Database Reference INTERPRO; IPR002624; Comment: This family consists of various deoxynucleoside kinases Comment: cytidine EC:2.7.1.74, guanosine EC:2.7.1.113, adenosine EC:2.7.1.76 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyuridine Comment: and deoxycytosine.) These enzymes catalyse the production of Comment: deoxynucleotide 5'-monophosphate from a deoxynucleoside. Comment: Using ATP and yielding ADP in the process. Number of members: 20
DSL		Delta serrate ligand	Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96125168 Reference Title: Interchangeability of Caenorhabditis elegans DSL proteins Reference Title: and intrinsic signalling activity of their extracellular Reference Title: domains in vivo. Reference Author: Fitzgerald K, Greenwald I; Reference Location: Development 1995;121:4275-4282. Reference Number: [2] Reference Medline: 92034990 Reference Title: Specific EGF repeats of Notch mediate interactions with Reference Title: Delta and Serrate: implications for Notch as a Reference Title: multifunctional receptor. Reference Author: Rebay I, Fleming RJ, Fehon RG, Cherbas L, Cherbas P, Reference Author: Artavanis-Tsakonas S; Reference Location: Cell 1991;67:687-699. Reference Number: [3] Reference Medline: 95232495 Reference Title: Notch signaling. Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30
DUF125		Integral membrane protein DUF125	Accession number: PF01988 Definition: Integral membrane protein DUF125

Pfam	Prosite	Full Name	Description
			Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -60 -60 Trusted cutoffs: -57.90 -57.90 Noise cutoffs: -64.60 -64.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95028150 Reference Title: Sequence, mapping and disruption of CCC1, a gene that Reference Title: cross-complements the Ca(2+)-sensitive phenotype of csg1 Reference Title: mutants. Reference Author: Fu D, Beeler T, Dunn T; Reference Location: Yeast 1994;10:515-521. Database Reference INTERPRO; IPR002839; Comment: This family of predicted integral membrane proteins has no known Comment: function. However it does include Swiss:P47818, that may have a Comment: role in regulating calcium levels [1]. Number of members: 7
DUF25		Domain of unknown function DUF25	Accession number: PF01641 Definition: Domain of unknown function DUF25 Author: Bateman A, Enwright A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1539 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 151.80 151.80 Noise cutoffs: 10.60 10.60 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 20076492 Reference Title: Novel selenoproteins identified in silico and in vivo by Reference Title: using a conserved RNA structural motif. Reference Author: Lescure A, Gautheret D, Carbon P, Krol A; Reference Location: J Biol Chem 1999;274:38147-38154. Database Reference INTERPRO; IPR002579; Comment: This domain has no known function. It is found associated Comment: with the peptide methionine sulfoxide reductase enzymatic Comment: domain PMSR. The domain has two conserved cysteine Comment: and histidines that could suggest and zinc binding site. Comment: The final cysteine is found to be replaced by the rare amino Comment: acid selenocysteine in some members of the family [1]. Number of members: 26
DUF26		Domain of unknown function DUF26	Accession number: PF01657 Definition: Domain of unknown function DUF26 Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_980 (release 4.1) Gathering cutoffs: -8 -8 Trusted cutoffs: 6.50 1.40 Noise cutoffs: -17.50 -17.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002902; Database reference: PFAMB; PB005223; Comment: This domain has no known function. It is found in serine/threonine Comment: kinases, associated with the Eukaryotic protein kinase domain

Pfam	Prosite	Full Name	Description
			Comment: pkinase. In the 33kDa secretary protein Swiss:082551 Comment: this domain is duplicated. The domain contains four conserved Comment: cysteines. Number of members: 25
DUF89			Accession number: PF01937 Definition: Protein of unknown function DUF89 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 636.30 636.30 Noise cutoffs: -142.40 -142.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002791; Comment: This prokaryotic family has no known function. The protein Comment: has two closely spaced conserved cysteines at its N Comment: terminus and a single conserved cysteine at its C terminus. Number of members: 5
DUF90			Accession number: PF01938 Definition: Domain of unknown function DUF90 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs. 25 0 Trusted cutoffs: 78.90 10.20 Noise cutoffs: -0.60 -0.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002792; Comment: This small domain has no known function. However it Comment: may perform a nucleic acid binding role (Bateman A. Comment: unpublished observation). Number of members: 17
Dynein_light	PDOC00953	Dynein light chain type 1 signature	Dynein is a multisubunit microtubule-dependent motor enzyme that acts as the force generating protein of eukaryotic cilia and flagella. The cytoplasmic isoform of dynein acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules. Dynein is composed of a number of ATP-binding large subunits, intermediate size subunits and small subunits. Among the small subunits, there is a family [1,2] of highly conserved proteins which consist of: - Chlamydomonas reinhardtii flagellar outer arm dynein 8 Kd and 11 Kd light chains. - Higher eukaryotes cytoplasmic dynein light chain 1. - Yeast cytoplasmic dynein light chain 1 (gene DYN2 or SLC1). - Caenorhabditis elegans hypothetical dynein light chains M18.2 and T26A5.9. These proteins are have from 89 to 120 amino acids. As a signature pattern, we selected a highly conserved region. Description of pattern(s) and/or profile(s)

Pfam	Prosite	Full Name	Description
			Consensus pattern H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / First entry. References [1] King S.M., Patel-King R.S. J. Biol. Chem. 270:11445-11452(1995). [2] Dick T., Ray K., Salz H.K., Chia W. Mol. Cell. Biol. 16:1966-1977(1996).
eIF5_eIF2B		Domain found in IF2B/IF5	Accession number: PF01873 Definition: Domain found in IF2B/IF5 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 233.00 233.00 Noise cutoffs: -56.10 -56.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96060092 Reference Title: Multidomain organization of eukaryotic guanine nucleotide Reference Title: exchange translation initiation factor eIF-2B subunits Reference Title: revealed by analysis of conserved sequence motifs. Reference Author: Koonin EV; Reference Location: Protein Sci 1995;4:1608-1617. Database Reference INTERPRO; IPR002735; Comment: This family includes the N terminus of eIF-5 Swiss:P55010, and Comment: the C terminus of eIF-2 beta Swiss:P20042. This region Comment: corresponds to the whole of the archaebacterial eIF-2 beta Comment: homolog. The region contains a putative zinc binding C4 finger. Number of members: 20
eIF6		eIF-6 family	This family comprises members exhibiting sequence identity to the eukaryotic translation initiation factor 6. Some members of this family are implicated in protein biosynthesis as a translation initiation factor by binding to the 60s ribosomal subunit and preventing its association with the 40s ribosomal subunit to form the 80s initiation complex. Such activity can play a role in maximal polysome formation and plays an important role in determining free 60s ribosomal subunit content. Polypeptides in this family can optimize amino acid and nitrogen content in a desired cell or organism. References describing eif6 family members and their biological activities include, for example, the following: Adams et al., Science 87:2185-2195(2000); Wood et al., J. Biol. Chem. 274:11653-11659(1999); and Si et al., Mol. Cell. Biol. 19:1416-1426(1999).
ER	PDOC00992	Enhancer of rudimentary signature	The Drosophila protein 'enhancer of rudimentary' (gene (e(r)) is a small protein of 104 residues whose function is not yet clear. From an evolutionary point of view, it is highly conserved [1] and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein plays a role in the cell cycle As as signaure pattern, we selected a conserved region in the central part of

Pfam	Prosite	Full Name	Description
			the protein. Description of pattern(s) and/or profile(s) Consensus pattern Y-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / First entry. References [1] Gelsthorpe M., Pulumati M., McCallum C., Dang-Vu K., Tsubota S.I. Gene 186:189-195(1997).
ER_lumen_recept	PDOC00732	ER lumen protein retaining receptor signatures	Proteins that reside in the lumen of the endoplasmic reticulum (ER) contain a C-terminal tetrapeptide (generally K-D-E-L or H-D-E-L) that serves as a signal for their retrieval (retrograde transport) from subsequent compartments of the secretory pathway. The signal is recognized by a receptor molecule that is believed to cycle between the cis side of the Golgi apparatus and the ER [1]. This protein is known as the ER lumen protein retaining receptor or also as the 'KDEL receptor'. It has been characterized in a variety of species, including fungi (gene ERD2), plants, Plasmodium, Drosophila and mammals. In mammals two highly related forms of the receptor are known. Structurally, the receptor is a protein of about 220 residues that seems to contain seven transmembrane regions [2]. The N-terminal part (3 residues) is oriented toward the lumen while the C-terminal tail (about 12 residues) is cytoplasmic. There are three lumenal and three cytoplasmic loops. We developed two signature patterns for these receptors. The first pattern corresponds to the C-terminal half of the first cytoplasmic loop as well as most of the second transmembrane domain. The second pattern is a perfectly conserved decapeptide that corresponds to the central part of the fifth transmembrane domain. Description of pattern(s) and/or profile(s) Consensus pattern G-[LIV]-S-x-[KR]-x-[QH]-x-L-[FY]-x-[LIV](2)-[FYW]-x(2)-R- Y Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern L-E-[SA]-V-A-I-[LM]-P-Q-[LI] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Patterns and text revised. References [1] Pelham H.R.B.

Pfam	Prosite	Full Name	Description
			Curr. Opin. Cell Biol. 3:585-591(1991). [2] Townsley F.M., Wilson D.W., Pelham H.R.B. EMBO J. 12:2821-2829(1993).
ETF_alpha	PDOC00583	Electron transfer flavoprotein alpha-subunit signature	The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria. The alpha subunit of ETF is a protein of about 32 Kd which is structurally related to the bacterial nitrogen fixation protein fixB which could play a role in a redox process and feed electrons to ferredoxin. Other related proteins are: - Escherichia coli hypothetical protein ydiR. - Escherichia coli hypothetical protein ygcQ. As a signature pattern for these proteins we selected a highly conserved region which is located in the C-terminal section. Description of pattern(s) and/or profile(s) Consensus pattern [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)- [IV]-x-A-[IV]-N Sequences known to belong to this class detected by the pattern ALL, except for ygcQ. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Text revised. References [1] Finocchiaro G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990). [2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).
Euk_porin	PDOC00483	Eukaryotic mitochondrial porin signature	The major protein of the outer mitochondrial membrane of eukaryotes is a porin that forms a voltage-dependent anion-selective channel (VDAC) that behaves as a general diffusion pore for small hydrophilic molecules [1 to 4]. The channel adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV. This protein contains about 280 amino acids and its sequence is composed of between 12 to 16 beta-strands that span the mitochondrial outer membrane. Yeast contains two members of this family (genes POR1 and POR2); vertebrates have at least three members (genes VDAC1, VDAC2 and VDAC3) [5]. As a signature pattern we selected a conserved region located at the C- terminal part of these proteins.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern [YH]-x(2)-D-[SPCAD]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]- [DNSTA]-[DNS]-x(4)-[GSTAN]-[LIVMA]-x-[LIVMY] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Benz R. Biochim. Biophys. Acta 1197:167-196(1994). [2] Manella C.A. Trends Biochem. Sci. 17:315-320(1992). [3] Dihanich M. Experientia 46:146-153(1990). [4] Forte M., Guy H.R., Mannella C.A. J. Bioenerg. Biomembr. 19:341-350(1987). [5] Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J. Genomics 36:192-196(1996).
F_bP_aldolase	PDOC00523	Fructose-bisphosphate aldolase class-II signatures	Fructose-bisphosphate aldolase (EC 4.1.2.13) [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6- bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate. There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-II aldolases [2], mainly found in prokaryotes and fungi, are homodimeric enzymes which require a divalent metal ion - generally zinc - for their activity. This family also includes the following proteins: - Escherichia coli galactitol operon protein gatY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D- glyceraldehyde 3-phosphate. - Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the same reaction as that of gatY. As signature patterns for this class of enzyme, we selected two conserved regions. The first pattern is located in the first half of the sequence and contains two histidine residues that have been shown [4] to be involved in binding a zinc ion. The second is located in the C-terminal section and contains clustered acidic residues and glycines. Description of pattern(s) and/or profile(s) Consensus pattern [FYVMT]-x(1,3)-[LIVMH]-[APNT]-[LIVM]-x(1,2)-[LIVM]-H-x-D- H-[GACH] [The two H's are zinc ligands]

Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL, except for Mycoplasma pneumoniae aldolase. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990). [2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992). [3] von der Osten C.H., Barbas C.F. III, Wong C.-H., Sinskey A.J. Mol. Microbiol. 3:1625-1637(1989). [4] Berry A., Marshall K.E. FEBS Lett. 318:11-16(1993).
FAA_hydrolase		Fumarylacetoacetate (FAA) hydrolase family	Accession number: PF01557 Definition: Fumarylacetoacetate (FAA) hydrolase family Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_641 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 42.10 42.10 Noise cutoffs: -93.10 -93.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97255958 Reference Title: Mutations in the fumarylacetoacetate hydrolase gene causing Reference Title: hereditary tyrosinemia type I: overview. Reference Author: St-Louis M, Tanguay RM; Reference Location: Hum Mutat 1997;9:291-299. Reference Number: [2] Reference Medline: 96125235 Reference Title: Molecular characterization of the 4-hydroxyphenylacetate Reference Title: catabolic pathway of Escherichia coli W: engineering a Reference Title: mobile aromatic degradative cluster. Reference Author: Prieto MA, Diaz E, Garcia JL; Reference Location: J Bacteriol 1996;178:111-120. Reference Number: [3] Reference Medline: 96016123 Reference Title: Fungal metabolic model for human type I hereditary Reference Title: tyrosinaemia. Reference Author: Fernandez-Canon JM, Penalva MA; Reference Location: Proc Natl Acad Sci U S A 1995;92:9132-9136. Reference Number: [4] Reference Medline: 94039092 Reference Title: Purification, nucleotide sequence and some properties of a Reference Title: bifunctional isomerase/decarboxylase from the Reference Title: homoprotocatechuate degradative pathway of Escherichia coli Reference Title: C. Reference Author: Roper DI, Cooper RA; Reference Location: Eur J Biochem 1993;217:575-580. Database reference: MIM; 276700; Database Reference INTERPRO; IPR002529;

Pfam	Prosite	Full Name	Description
			Comment: This family consists of fumarylacetoacetate (FAA) hydrolase, Comment: or fumarylacetoacetate hydrolase (FAH) and it also includes Comment: HHDD isomerase/OPET decarboxylase from E. coli strain W. Comment: FAA is the last enzyme in the tyrosine catabolic pathway, it hydrolyses Comment: fumarylacetoacetate into fumarate and acetoacetate which then join the Comment: citric acid cycle [1]. Mutations in FAA cause type I tyrosinemia in humans Comment: this is an inherited disorder mainly affecting the liver leading to Comment: liver cirrhosis, hetpatocellular carcinoma, renal tubular damages and Comment: neurologic crises amongst other symptoms [1]. The enzymatic defect causes Comment: the toxic accumulation of phenylalanine/tyrosine catabolites [3]. Comment: The E. coli W enzyme HHDD isomerase/OPET decarboxylase contains two Comment: copies of this domain and functions in fourth and fifth steps of the Comment: homoprotocatechuate pathway; Comment: here it decarboxylates OPET to HHDD and isomerizes this to OHED. Comment: The final products of this pathway are pyruvic acid and succinic Comment: semialdehyde. Number of members: 33
FAD_binding		FAD binding domain	Accession number: PF00667 Definition: FAD binding domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_180 (release 2.1) Gathering cutoffs: 16.8 16.8 Trusted cutoffs: 24.60 16.80 Noise cutoffs: 13.50 15.90 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95386502 Reference Title: The flavin reductase activity of the flavoprotein component Reference Title: of sulfite reductase from Escherichia coli. A new model for Reference Title: the protein structure. Reference Author: Eschenbrenner M, Coves J, Fontecave M; Reference Location: J Biol Chem 1995;270:20550-20555. Reference Number: [2] Reference Medline: 96049560 Reference Title: NADPH-sulfite reductase flavoprotein from Escherichia coli: Reference Title: contribution to the flavin content and subunit interaction. Reference Author: Eschenbrenner M, Coves J, Fontecave M; Reference Location: FEBS Lett 1995;374:82-84. Reference Number: [3] Reference Medline: 94360001 Reference Title: Dissection of NADPH-cytochrome P450 oxidoreductase into Reference Title: distinct functional domains. Reference Author: Smith GC, Tew DG, Wolf CR; Reference Location: Proc Natl Acad Sci U S A 1994;91:8710-8714. Reference Number: [4] Reference Medline: 97385116 Reference Title: Three-dimensional structure of NADPH-cytochrome P450 Reference Title: reductase: prototype for FMN- and FAD-containing enzymes. Reference Author: Wang M, Roberts DL, Paschke R, Shea

Pfam	Prosite	Full Name	Description
			TM, Masters BS, Kim JJ; Reference Location: Proc Natl Acad Sci U S A 1997;94:8411-8416. Database Reference: SCOP; 1amo; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001709; Database Reference PDB; 1amo A; 274; 493; Database Reference PDB; 1amo B; 274; 493; Database Reference PDB; 1quf ; 77; 120; Database reference: PFAMB; PB001390; Comment: This domain is found in sulfite reductase, NADPH cytochrome P450 Comment: reductase and Nitric oxide synthase. Number of members: 87
FAD_binding_3		FAD binding domain	Accession number: PF01494 Definition: FAD binding domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_549 (release 4.0) Gathering cutoffs: -7 -7 Trusted cutoffs: -6.20 -6.20 Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93028353 Reference Title: Crystal structure of the reduced form of p-hydroxybenzoate Reference Title: hydroxylase refined at 2.3A resolution. Reference Author: Schreuder HA, van der Laan JM, Swarte MB, Kalk KH, Hol WG, Reference Author: Drenth J; Reference Location: Proteins 1992;14:178-190. Database Reference: SCOP; 2phh; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002938; Database Reference PDB; 1pxa ; 5; 35; Database Reference PDB; 1bf3 ; 5; 139; Database Reference PDB; 1bgj ; 5; 139; Database Reference PDB; 1bgn ; 5; 139; Database Reference PDB; 1bkw ; 5; 139; Database Reference PDB; 1cc4 A; 5; 139; Database Reference PDB; 1cc6 A; 5; 139; Database Reference PDB; 1cj2 A; 5; 139; Database Reference PDB; 1pbb ; 5; 139; Database Reference PDB; 1pbc ; 5; 139; Database Reference PDB; 1pbd ; 5; 139; Database Reference PDB; 1pbe ; 5; 139; Database Reference PDB; 1pbf ; 5; 139; Database Reference PDB; 1pdh ; 5; 139; Database Reference PDB; 2phh ; 5; 139; Database Reference PDB; 1cj3 A; 5; 139; Database Reference PDB; 1cj4 A; 5; 139; Database Reference PDB; 1phh ; 5; 139; Database Reference PDB; 1d7l A; 5; 139; Database Reference PDB; 1dob ; 5; 139; Database Reference PDB; 1doc ; 5; 139; Database Reference PDB; 1dod ; 5; 139; Database Reference PDB; 1doe ; 5; 139; Database Reference PDB; 1ius ; 5; 139; Database Reference PDB; 1iut ; 5; 139; Database Reference PDB; 1iuu ; 5; 139; Database Reference PDB; 1iuv ; 5; 139; Database Reference PDB; 1iuw ; 5; 139; Database Reference PDB; 1iux ; 5; 139; Database Reference PDB; 1foh A; 10; 151; Database Reference PDB; 1foh D; 10; 151; Database Reference PDB; 1foh B; 10; 151; Database Reference PDB; 1foh C; 10; 151; Database reference: PFAMB; PB040546; Comment: This domain is involved in FAD binding in a number of enzymes. Number of members: 52

Pfam	Prosite	Full Name	Description
FAD_binding_4	PDOC00674	Oxygen oxidoreductases covalent FAD-binding site	Some oxygen-dependent oxidoreductases are flavoproteins that contains a covalently bound FAD group which is attached to a histidine via an 8-alpha- (N3-histidyl)-riboflavin linkage. These proteins are: - 6-hydroxy-D-nicotine oxidase (EC 1.5.3.6) (6-HDNO) [1], a bacterial enzyme that catalyzes the oxygen-dependent degradation of 6-hydroxynicotine into 6-hydroxypyrid-N-methylosmine - Plant reticuline oxidase (EC 1.5.3.9) [2] (berberine-bridge-forming enzyme), an enzyme that catalyzes the oxidation of (S)-reticuline into (S)- scoulerine in the pathway leading to benzophenanthridine alkaloids. - L-gulonolactone oxidase (EC 1.1.3.8) (l-gulono-gamma-lactone oxidase) [3], a mammalian enzyme which catalyzes the oxidation of L-gulono-1,4-lactone to L-xylo-hexulonolactone which spontaneously isomerizes to L-ascorbate. - D-arabinono-1,4-lactone oxidase (EC 1.1.3.24) (L-galactonolactone oxidase), a yeast enzyme involved in the biosynthesis of D-erythroascorbic acid [4]. - Mitomycin radical oxidase [5], a bacterial protein involved in mitomycin resistance and that probably oxidizes the reduced form of mitomycins - Rhodococcus fascians fasciation locus protein fas5. The region around the histidine that binds the FAD group is conserved in these enzymes and can be used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern P-x(10)-[DE]-[LIVM]-x(3)-[LIVM]-x(9)-[LIVM]-x(3)-[GSA]- [GST]-G-H [H is the FAD binding site] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Text revised. EMBL/GenBank: U40390. References [1] Brandsch R., Hinkkanen A.E., Mauch L., Nagursky H., Decker K. Eur. J. Biochem. 167:315-320(1987). [2] Dittrich H., Kutchan T.M. Proc. Natl. Acad. Sci. U.S.A. 88:9969-9973(1991). [3] Koshizaka T., Nishikimi M., Ozawa T., Yagi K. J. Biol. Chem. 263:1619-1621(1988). [4] Huh W.-K., Kim S.-T., Kim J.-Y., Hwang S.-W., Kang S.-O. [5] August P.R., Flickinger M.C., Sherman D.H. J. Bacteriol. 176:4448-4454(1994).
fer2	PDOC00175; PDOC00642	2Fe-2S ferredoxins, iron-sulfur binding region signature; Adrenodoxin family, iron-sulfur binding	Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided

Pfam	Prosite	Full Name	Description
		region signature	into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One of these subgroups are the 2Fe-2S ferredoxins, which are proteins or domains of around one hundred amino acid residues that bind a single 2Fe-2S iron-sulfur cluster. The proteins that are known [2] to belong to this family are listed below. - Ferredoxin from photosynthetic organisms; namely plants and algae where it is located in the chloroplast or cyanelle; and cyanobacteria. - Ferredoxin from archaebacteria of the Halobacterium genus. - Ferredoxin IV (gene pftA) and V (gene fdxD) from Rhodobacter capsulatus. - Ferredoxin in the toluene degradation operon (gene xylT) and naphthalene degradation operon (gene nahT) of Pseudomonas putida. - Hypothetical Escherichia coli protein yfaE. - The N-terminal domain of the bifunctional ferredoxin/ferredoxin reductase electron transfer component of the benzoate 1,2-dioxygenase complex (gene benC) from Acinetobacter calcoaceticus, the toluene 4-monooxygenase complex (gene tmoF), the toluate 1,2-dioxygenase system (gene xylZ), and the xylene monooxygenase system (gene xylA) from Pseudomonas. - The N-terminal domain of phenol hydroxylase protein p5 (gene dmpP) from Pseudomonas Putida. - The N-terminal domain of methane monooxygenase component C (gene mmoC) from Methylococcus capsulatus . - The C-terminal domain of the vanillate degradation pathway protein vanB in a Pseudomonas species. - The N-terminal domain of bacterial fumarate reductase iron-sulfur protein (gene frdB). - The N-terminal domain of CDP-6-deoxy-3,4-glucoseen reductase (gene ascD) from Yersinia pseudotuberculosis. - The central domain of eukaryotic succinate dehydrogenase (ubiquinone) iron- sulfur protein. - The N-terminal domain of eukaryotic xanthine dehydrogenase. - The N-terminal domain of eukaryotic aldehyde oxidase. In the 2Fe-2S ferredoxins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region. Description of pattern(s) and/or profile(s) Consensus pattern C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C [The three C's are 2Fe-2S ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 15. Note in addition to the proteins listed above there are a number of other ferredoxin-like proteins that bind a 2Fe-2S cluster but which do not seem to be evolutionary related to this family. Among them are the ferredoxins from the adrenodoxin family (see <PDOC00642>) as well as the bacterial aromatic dioxygenase

Pfam	Prosite	Full Name	Description
			systems ferredoxin-like proteins such as bnzC, ndoA, and todB. Last update November 1997 / Text revised. References [1] Meyer J. Trends Ecol. Evol. 3:222-226(1988). [2] Harayama S., Polissi A., Rekik M. FEBS Lett. 285:85-88(1991). Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One family of ferredoxins groups together the following proteins that all bind a single 2Fe- 2S iron-sulfur cluster: - Adrenodoxin (ADX) (adrenal ferredoxin), a vertebrate mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450scc, which is involved in cholesterol side chain cleavage. - Putidaredoxin (PTX), a Pseudomonas putida protein which transfers electrons from putidaredoxin reductase to cytochrome P450-cam, which is involved in the oxidation of camphor. - Terpredoxin [2], a Pseudomonas protein which transfers electrons from terpredoxin reductase to cytochrome P450-terp, which is involved in the oxidation of alpha-terpineol. - Rhodocoxin [3], a Rhodococcus protein which transfers electrons from rhodocoxin reductase to cytochrome CYP116 (thcB), which is involved in the degradation of thiocarbamate herbicides. - Escherichia coli ferredoxin (gene fdx) [4] whose exact function is not yet known. - Rhodobacter capsulatus ferredoxin VI [5], which may transfer electrons to a yet uncharacterized oxygenase. - Caulobacter crescentus ferredoxin (gene fdxB) [6]. In these proteins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region. Description of pattern(s) and/or profile(s) Consensus pattern C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR] [The three C's are 2Fe-2S ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 1. Last update November 1995 / Pattern and text revised. EMBL/Genbank: X51607. References [1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).

Pfam	Prosite	Full Name	Description
			[2] Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S., Carmona C., Witney F., Lorence M.C. J. Biol. Chem. 267:14193-14203(1992). [3] Nagy I., Schoofs G., Compernolle F., Proost P., Vanderleyden J., De Mot R. J. Bacteriol. 177:676-687(1995). [4] Ta D.T., Vickery L.E. J. Biol. Chem. 267:11120-11125(1992). [5] Naud I., Vincon M., Garin J., Gaillard J., Forest E., Jouanneau Y. Eur. J. Biochem. 222:933-939(1994). [6] Amemiya K
Ferric_reduct		Ferric reductase like transmembrane component	Accession number: PF01794 Definition: Ferric reductase like transmembrane component Author: Bashton M, Bateman A Alignment method of seed: T_Coffee Source of seed members: Pfam-B_728 (release 4.2) Gathering cutoffs: -122 -122 Trusted cutoffs: -34.80 -34.80 Noise cutoffs: -210.30 -210.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93309468 Reference Title: The fission yeast ferric reductase gene frp1+ is required Reference Title: for ferric iron uptake and encodes a protein that is Reference Title: homologous to the gp91-phox subunit of the human NADPH Reference Title: phagocyte oxidoreductase. Reference Author: Roman DG, Dancis A, Anderson GJ, Klausner RD; Reference Location: Mol Cell Biol 1993;13:4342-4350. Reference Number: [2] Reference Medline: 92294876 Reference Title: Cytochrome b558: the flavin-binding component of the Reference Title: phagocyte NADPH oxidase. Reference Author: Rotrosen D, Yeung CL, Leto TL, Malech HL, Kwong CH; Reference Location: Science 1992;256:1459-1462. Reference Number: [3] Reference Medline: 87258189 Reference Title: The glycoprotein encoded by the X-linked chronic Reference Title: granulomatous disease locus is a component of the Reference Title: neutrophil cytochrome b complex. Reference Author: Dinauer MC, Orkin SH, Brown R, Jesaitis AJ, Parkos CA; Reference Location: Nature 1987;327:717-720. Reference Number: [4] Reference Medline: 87258190 Reference Title: The X-linked chronic granulomatous disease gene codes for Reference Title: the beta- chain of cytochrome b-245. Reference Author: Teahan C, Rowe P, Parker P, Totty N, Segal AW; Reference Location: Nature 1987;327:720-721. Database Reference INTERPRO; IPR002916; Comment: This family includes a common region in the transmembrane proteins Comment: mammalian cytochrome B-245 heavy chain

Pfam	Prosite	Full Name	Description
			(gp91-phox), ferric reductase Comment: transmembrane component in yeast and respiratory burst oxidase from Comment: mouse-ear cress. Comment: This may be a family of flavocytochromes capable of moving electrons Comment: across the plasma membrane [1]. Comment: The Frp1 protein Swiss:Q04800 from S. pombe is a ferric reductase Comment: component and is required for cell surface ferric reductase activity, Comment: mutants in frp1 are deficient in ferric iron uptake [1]. Comment: Cytochrome B-245 heavy chain Swiss:P04839 is a FAD-dependent Comment: dehydrogenase it is also has electron transferase activity which reduces Comment: molecular oxygen to superoxide anion, a precursor in the production of Comment: microbicidal oxidants [2]. Comment: Mutations in the sequence of cytochrome B-245 heavy chain (gp91-phox) Comment: lead to the X-linked chronic granulomatous disease. The bacteriocidal Comment: ability of phagocytic cells is reduced and is characterised by the Comment: absence of a functional plasma membrane associated NADPH oxidase [3]. Comment: The chronic granulomatous disease gene codes for the beta chain of Comment: cytochrome B-245 and cytochrome B-245 is missing from patients with Comment: the disease [4]. Comment: The aligned region includes a potential FAD binding domain. Number of members: 34
Flavi_NS5		Flavivirus RNA-directed RNA polymerase	Accession number: PF00972 Definition: Flavivirus RNA-directed RNA polymerase Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_200 (release 3.0) Gathering cutoffs: 12 12 Trusted cutoffs: 16.00 16.00 Noise cutoffs: 8.50 8.50 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95159427 Reference Title: Phylogeny of TYU, SRE, and CFA virus: different Reference Title: evolutionary rates in the genus Flavivirus. Reference Author: Marin MS, Zanotto PM, Gritsun TS, Gould EA; Reference Location: Virology 1995;206:1133-1139. Reference Number: [2] Reference Medline: 96182933 Reference Title: Recombinant dengue type 1 virus NS5 protein expressed in Reference Title: Escherichia coli exhibits RNA-dependent RNA polymerase Reference Title: activity. Reference Author: Tan BH, Fu J, Sugrue RJ, Yap EH, Chan YC, Tan YH; Reference Location: Virology 1996;216:317-325. Reference Number: [3] Reference Medline: 93224895 Reference Title. Computer-assisted identification of a putative Reference Title: methyltransferase domain in NS5 protein of flaviviruses and Reference Title: lambda 2 protein of reovirus. Reference Author: Koonin EV; Reference Location: J Gen Virol 1993;74:733-740.

Pfam	Prosite	Full Name	Description
			Reference Number: [4] Reference Medline: 94094568 Reference Title: Evolution and taxonomy of positive-strand RNA viruses: Reference Title: implications of comparative analysis of amino acid Reference Title: sequences. Reference Author: Koonin EV, Dolja VV; Reference Location: Crit Rev Biochem Mol Biol 1993;28:375-430. Database Reference INTERPRO; IPR000208; Comment: Flaviviruses produce a polyprotein from the ssRNA genome. Comment: This protein is also known as NS5. Comment: This RNA-directed RNA polymerase possesses a number of short Comment: regions and motifs homologous to other RNA-directed RNA Comment: polymerases [2]. Number of members: 159
Fork_head	PDOC00564	Fork head domain signatures and profile	It has been shown [1] that some eukaryotic transcription factors contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below. - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5. - Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericim-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia. - Human FKHR which is involved in a chromosomal

Pfam	Prosite	Full Name	Description
			translocation that causes rhabdomyosarcoma. - Xenopus XFKH1, a protein essential for normal axis formation. - Caenorhabditis elegans lin-31; involved in the regulation of vulval cell fates. - Yeast HCM1, a protein of unknown function. - Yeast FKH1. - Yeast FKH2. The fork domain is highly conserved. We have developed two patterns for its detection. The first corresponds to the N-terminal section of the domain; the second is a heptapeptide located in the central section of the domain. Description of pattern(s) and/or profile(s) Consensus pattern [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]- [LIM] Sequences known to belong to this class detected by the pattern ALL, except for AFX1 and FKHR. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern W-[QKR]-[NS]-S-[LIV]-R-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Weigel D., Jaeckle H. Cell 63:455-456(1990). [2] Clark K.L., Halay E.D., Lai E., Burley S.K. Nature 364:412-420(1993). [3] Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel W., Jaeckle H. EMBO J. 14:5306-5317(1995).
FtsJ		FtsJ cell division protein	Accession number: PF01728 Definition: FtsJ cell division protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1791 (release 4.1) Gathering cutoffs: -38 -38 Trusted cutoffs: -20.90 -20.90 Noise cutoffs: -56.70 -56.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93186701 Reference Title: The Escherichia coli FtsH protein is a prokaryotic member Reference Title: of a protein family of putative ATPases involved in Reference Title: membrane functions, cell cycle control, and gene Reference Title: expression. Reference Author: Tomoyasu T, Yuki T, Morimura S, Mori H, Yamanaka K, Niki H, Reference Author: Hiraga S, Ogura T; Reference Location: J Bacteriol 1993;175:1344-1351. Database Reference INTERPRO; IPR002877; Database reference: PFAMB; PB030182; Comment: This family consists of FtsJ from various bacterial and archaeal sources

Pfam	Prosite	Full Name	Description
			Comment: In E. coli FtsJ is not essential for growth but affects cell division [1]. Number of members: 25
FTSW RODA SPO VE	PDOC00352	Cell cycle proteins ftsW / rodA / spoVE signature	A number of prokaryotic proteins involved in cell cycle processes have been found [1,2] to be structurally related. these proteins are: - Escherichia coli and related bacteria cell division protein ftsW. This protein plays a role in the stabilization of the ftsZ ring during cell division. - Escherichia coli and related bacteria rod shape-determining protein rodA (or mrdB). It is required for the expression of the enzymatic activity of PBP2, which is thought to participate in the synthesis of peptidoglycan during the initiation of cell elongation. - Bacillus subtilis stage V sporulation protein E (spoVE). The exact function of spoVE in endospore formation is not known. - Bacillus subtilis hypothetical protein ylaO. - Bacillus subtilis hypothetical protein ywcF (ipa-42D). - Cyanophora paradoxa cyanelle ftsW homolog. This protein may be involved in the organelle division process. All these proteins are hydrophobic integral membrane protein and contain about 400 residues. We have selected the best conserved region, which is located in the C-terminal section, as a signature pattern for these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)- [LIVMFW](2)-S-[YSA]-G-G-[STN]-[SA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y., Matsuhashi M. J. Bacteriol. 171:6375-6378(1989). [2] Joris B., Dive G., Henriques A., Piggot P.J., Ghuysen J.-M. Mol. Microbiol. 4:513-517(1990).
Furin-like		Furin-like cysteine rich region	Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein. References describing these family members and their biological activities:

Pfam	Prosite	Full Name	Description
			Abbot et al., J. Biol. Chem. 267:10759-10763(1992);Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum Mol. Genet. 2:1437-1441(1993); Clifford et al., Genetics 137:531-550(1994); Cocozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al., Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Proc. Natl. Acad. Sci. U.S.A. 84:704-708(1987); Ehsani et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997). Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 266:5260-5267(1991); Lai et al., Neuron 6:691-704(1991); Lax et al., Mol. Cell. Biol. 8:1970-1978(1988); Lebrun et al., J. Biol. Chem. 268:11272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985). Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Mol. Endocrinol. 4:1183-1191(1990); Odawara et al., Science 245:66-68(1989); Raz et al., Genetics 129:191-201(1991). Sakai et al., J. Mol. Biol. 256:548-555(1996); Schaeffer et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Cell 46:1091-1101(1986); Seino et al., Biochem. Biophys. Res. Commun. 159:312-316(1989); Seino et al., Diabetes 39:123-128(1990); Semba et al., Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985); Shier et al., J. Biol. Chem. 264:14605-14608(1989); Taira et al., Science 245:63-66(1989); Tewari et al., J. Biol. Chem. 264:16238-16245(1989); Ullrich et al., Nature 313:756-761(1985). Ullrich et al., EMBO J. 5:2503-2512(1986); van der Vorm et al., Diabetologia 36:172-174(1993); van der Vorm et al., J. Biol. Chem. 267:66-71(1992); Wadsworth et al., Nature 314:178-180(1985); White et al., Cell 54:641-649(1988); Xu et al., J. Biol. Chem. 265:18673-18681(1990); Yamamoto et al., Nature 319:230-234(1986); and Yoshimasa et al., Science 240:784-787(1988).
Galactosyl_T		Galactosyltransferase	Accession number: PF01762 Definition: Galactosyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_885 (release 4.2) Gathering cutoffs: -46 -46 Trusted cutoffs: -43.90 -43.90 Noise cutoffs: -49.80 -49.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98079080 Reference Title: Cloning of a human Reference Title: UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta- Reference Title: galactosyltransferase catalyzing the formation of type 1 Reference Title: chains. Reference Author: Kolbinger F, Streiff MB, Katopodis AG; Reference Location: J Biol Chem 1998;273:433-440. Reference Number: [2] Reference Medline: 98079027 Reference Title: Genomic cloning and expression of three

Pfam	Prosite	Full Name	Description
			murine Reference Title: UDP-galactose: beta-N- acetylglucosamine Reference Title: beta1,3-galactosyltransferase genes. Reference Author: Hennet T, Dinter A, Kuhnert P, Mattu TS, Rudd PM, Berger Reference Author: EG; Reference Location: J Biol Chem 1998;273:58-65. Database Reference INTERPRO; IPR002659; Database reference: PFAMB; PB005938; Database reference: PFAMB; PB012965; Comment: This family includes the galactosyltransferases Comment: UDP-galactose:2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase Comment: Swiss:O43825 [1] and UDP-Gal:beta-GlcNAc beta 1,3-galactosyltranferase Comment: Swiss:O54904 [2]. Comment: Specific galactosyltransferases transfer galactose to GlcNAc terminal Comment: chains in the synthesis of the lacto-series oligosaccharides types 1 Comment: and 2 [1]. Number of members: 29
G-alpha		G-protein alpha subunit	Accession number: PF00503 Definition: G-protein alpha subunit Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_11 (release 1.0) Gathering cutoffs: 13.8 13.8 Trusted cutoffs: 13.80 13.80 Noise cutoffs: 9.70 12.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94353239 Reference Title: Structures of active conformations of Gi alpha 1 and the Reference Title: mechanism of GTP hydrolysis. Reference Author: Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Reference Author: Sprang SR; Reference Location: Science 1994;265:1405-1412. Reference Number: [2] Reference Medline: 97004345 Reference Title: How G proteins work: a continuing story. Reference Author: Coleman DE, Sprang SR; Reference Location: Trends Biochem Sci 1996;21:41-44. Database Reference: PRINTS; PR00318; Database Reference: SCOP; 1gia; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001019; Database Reference PDB; 1gia ; 34; 343; Database Reference PDB; 1gil ; 34; 343; Database Reference PDB; 1as0 ; 32; 344; Database Reference PDB; 1gfi ; 33; 345; Database Reference PDB; 1as2 ; 32; 346; Database Reference PDB; 1bh2 ; 32; 346; Database Reference PDB; 1cip A; 32; 347; Database Reference PDB; 1git ; 32; 348; Database Reference PDB; 1agr D; 11; 353; Database Reference PDB; 1gg2 A; 6; 348; Database Reference PDB; 1gp2 A; 6; 348; Database Reference PDB; 1bof ; 10; 353; Database Reference PDB; 1as3 ; 9; 353; Database Reference PDB; 1gdd ; 9; 353; Database Reference PDB; 1agr A; 6; 353; Database Reference PDB; 1tag ; 27; 340; Database Reference PDB; 1tad A; 27; 342; Database Reference PDB; 1tad B; 27; 342; Database Reference PDB; 1tnd B; 27; 342; Database Reference PDB; 1tnd C; 27; 342; Database Reference PDB; 1tad C; 27; 344; Database Reference PDB; 1tnd A; 27; 349;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1cjk C; 39; 388; Database Reference PDB; 1cjt C; 39; 388; Database Reference PDB; 1cju C; 39; 388; Database Reference PDB; 1cjv C; 39; 388; Database Reference PDB; 1azt A; 35; 391; Database Reference PDB; 1azt B; 35; 391; Database Reference PDB; 1azs C; 36; 393; Database reference: PFAMB; PB034080; Comment: G proteins couple receptors of extracellular signals to intracellular Comment: signaling pathways. Comment: The G protein alpha subunit binds guanyl nucleotide and is a weak Comment: GTPase. Number of members: 245
GCV_H		Glycine cleavage H-protein	Accession number: PF01597 Definition: Glycine cleavage H-protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_988 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 27.90 27.90 Noise cutoffs: -58.80 -58.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94255425 Reference Title: X-ray structure determination at 2.6-A resolution of a Reference Title: lipoate- containing protein: the H-protein of the glycine Reference Title: decarboxylase complex from pea leaves. Reference Author: Pares S, Cohen-Addad C, Sieker L, Neuburger M, Douce R; Reference Location: Proc Natl Acad Sci U S A 1994;91:4850-4853. Database Reference: SCOP; 1htp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002930; Database Reference PDB; 1hpc A; 2; 127; Database Reference PDB; 1hpc B; 2; 127; Database Reference PDB; 1htp ; 2; 127; Comment: This is a family of glycine cleavage H-proteins, part of the glycine Comment: cleavage multienzyme complex (GCV) found in bacteria and the mitochondria Comment: of eukaryotes. GCV catalyses the catabolism of glycine in eukaryotes. Comment: A lipoyl group is attached to a completely conserved lysine residue. Comment: The H protein shuttles the methylamine group of glycine from the Comment: P protein to the T protein. Number of members: 40
GCV_T		Glycine cleavage T-protein (aminomethyl transferase)	Accession number: PF01571 Definition: Glycine cleavage T-protein (aminomethyl transferase) Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_933 (release 4.0) Gathering cutoffs: -146 -146 Trusted cutoffs: -124.50 -124.50 Noise cutoffs: -167.90 -167.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97199363 Reference Title: Cloning, and molecular characterization of the GCV1 gene Reference Title: encoding the glycine cleavage T-protein from Saccharomyces Reference Title: cerevisiae.

Pfam	Prosite	Full Name	Description
			Reference Author: McNeil JB, Zhang F, Taylor BV, Sinclair DA, Pearlman RE, Reference Author: Bognar AL; Reference Location: Gene 1997;186:13-20. Database Reference INTERPRO; IPR002536; Database reference: PFAMB; PB004229; Comment: This is a family of glycine cleavage T-proteins, part of the glycine Comment: cleavage multienzyme complex (GCV) found in bacteria and the mitochondria Comment: of eukaryotes. GCV catalyses the catabolism of glycine in eukaryotes. Comment: The T-protein is an aminomethyl transferase. Number of members: 27
G-gamma	PDOC01002	G-protein gamma subunit profile	Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl- geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits. The Caenorhabditis elegans protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain. We have developed a profile that spans the complete length of the gamma subunit. Description of pattern(s) and/or profile(s) Sequences known to belong to this class detected by the profile ALL, except for yeast and squid G-protein gamma. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Pennington S.R. srpenn@liverpool.ac.uk Last update November 1997 / First entry. References [1] Pennington S.R. Protein Prof. 2:16-315(1995).
glutaredoxin	PDOC00173	Glutaredoxin	Glutaredoxin [1,2,3], also known as thioltransferase, is a small protein of approximately one hundred amino-acid residues. It functions as an electron carrier in the glutathione-dependent synthesis of deoxyribonucleotides by the enzyme ribonucleotide reductase. Like thioredoxin, which functions in a similar way, glutaredoxin possesses an active center disulfide bond. It exists in either a reduced or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond.

Pfam	Prosite	Full Name	Description
			Glutaredoxin has been sequenced in a variety of species. On the basis of extensive sequence similarity, it has been proposed [4] that vaccinia protein O2L is most probably a glutaredoxin. Finally, it must be noted that phage T4 thioredoxin seems also to be evolutionary related. Description of pattern(s) and/or profile(s) Consensus pattern [LIVD]-[FYSA]-x(4)-C-[PV]-[FYWH]-C-x(2)-[TAV]-x(2,3)-[LIV] [The two C's form the redox-active bond] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note in position 5 of the pattern, all glutaredoxin sequences have Pro while T4 thioredoxin has Val. Last update December 1999 / Pattern and text revised. References [1] Gleason F.K., Holmgren A. FEMS Microbiol. Rev. 54:271-298(1988). [2] Holmgren A. Biochem. Soc. Trans. 16:95-96(1988). [3] Holmgren A. J. Biol. Chem. 264:13963-13966(1989). [4] Johnson G.P., Goebel S.J., Perkus M.E., Davis S.W., Winslow J.P., Paoletti E. Virology 181:378-381(1991).
Glyco_hydro_1	PDOC00495	Glycosyl hydrolases family 1 signatures	It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta-glucosidases (EC 3.2.1.21) from various bacteria such as Agrobacterium strain ATCC 21400, Bacillus polymyxa, and Caldocellum saccharolyticum. - Two plants (clover) beta-glucosidases (EC 3.2.1.21). - Two different beta-galactosidases (EC 3.2.1.23) from the archaebacteria Sulfolobus solfataricus (genes bgaS and lacS). - 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as Lactobacillus casei, Lactococcus lactis, and Staphylococcus aureus. - 6-phospho-beta-glucosidases (EC 3.2.1.86) from Escherichia coli (genes bglB and ascB) and from Erwinia chrysanthemi (gene arbB). - Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase). - Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62). LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is evolutionary related to the above glycosyl hydrolases. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-

Pfam	Prosite	Full Name	Description
			glucosidase from Agrobacterium, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. We have used this region as a signature pattern. As a second signature pattern we selected a conserved region, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue. Description of pattern(s) and/or profile(s) Consensus pattern [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 12. Note this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4]. Consensus pattern F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x- [GSTA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this pattern will pick up the last three domains of LPH. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1995 / Patterns and text revised. References [1] Henrissat B. Biochem. J. 280:309-316(1991). [2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991). [3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990). [4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992). [5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).
Glyco_hydro_19	PDOC00620	Chitinases family 19 signatures	Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19 (also known as classes IA or I and IB or II) are enzymes from plants that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the

Pfam	Prosite	Full Name	Description
			presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (see the relevant entry <PDOC00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues. As signature patterns we selected two highly conserved regions, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond. Description of pattern(s) and/or profile(s) Consensus pattern C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F- [GSA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Neuhaus J.-M. jean-marc.neuhaus@bota.unine.ch Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1997 / Text revised. References [1] Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992). [2] Henrissat B. Biochem. J. 280:309-316(1991). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt
Glyco_hydro_3_C	PDOC00621	Glycosyl hydrolases family 3 active site	It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta glucosidases (EC 3.2.1.21) from the fungi Aspergillus wentii (A-3), Hansenula anomala, Kluyveromyces fragilis. Saccharomycopsis fibuligera, (BGL1 and BGL2), Schizophyllum commune and Trichoderma reesei (BGL1). - Beta glucosidases from the bacteria Agrobacterium tumefaciens (Cbg1), Butyrivibrio fibrisolvens (bglA), Clostridium thermocellum (bglB), Escherichia coli (bglX), Erwinia chrysanthemi (bgxA) and Ruminococcus albus. - Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52). - Bacillus subtilis hypothetical protein yzbA. - Escherichica coli hypothetical protein ycfO and HI0959, the corresponding Haemophilus influenzae protein. One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus

Pfam	Prosite	Full Name	Description
			wentii beta- glucosidase A3, to be implicated in the catalytic mechanism. We have used this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]- [ST]-D-x(2)-[SGADNI] [D is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr Last update November 1997 / Pattern and text revised. References [1] Henrissat B. Biochem. J. 280:309-316(1991). [2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992). [3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).
Glyco_hydro_45	PDOC00877	Glycosyl hydrolases family 45 active site	The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family K or as the glycosyl hydrolases family 45 [3,E1]. The enzymes which are currently known to belong to this family are listed below. - Endoglucanase 5 from Humicola insolens. - Endoglucanase 5 from Trichoderma reesei (egl5). - Endoglucanase K from Fusarium oxysporum. - Endoglucanase B from Pseudomonas fluorescens (celB). - Endoglucanase 1 from Ustilago maydis (egl1). The best conserved regions in these enzymes is located in the N-terminal section. It contains an aspartic acid residue which has been shown [4] to act as a nucleophile in the catalytic mechanism. We use this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [STA]-T-R-Y-[FYW]-D-x(5)-[CA] [The D is an active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr

Pfam	Prosite	Full Name	Description
			Last update November 1997 / Pattern and text revised. References [1] Beguin P Annu. Rev. Microbiol. 44:219-248(1990). [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991). [3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993). [4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z., Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schuelein M. Nature 365:362-364(1993). [E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt
Glyco_hydro_47		Glycosyl hydrolase family 47	Members of this family are alpha-mannosidases that catalyse the hydrolysis of the terminal 1,2-linked alpha-D-mannose residues in the oligo-mannose oligosaccharide Man(9)(GlcNAc)(2). These enzymes are capable of taking part in the glycosylation pathway and glycoprotein processing.
GTP_cyclohydrol	PDOC00672	GTP cyclohydrolase I signatures	GTP cyclohydrolase I (EC 3.5.4.16) catalyzes the biosynthesis of formic acid and dihydroneopterin triphosphate from GTP. This reaction is the first step in the biosynthesis of tetrahydrofolate in prokaryotes, of tetrahydrobiopterin in vertebrates, and of pteridine-containing pigments in insects. GTP cyclohydrolase I is a protein of from 190 to 250 amino acid residues. The comparison of the sequence of the enzyme from bacterial and eukaryotic sources shows that the structure of this enzyme has been extremely well conserved throughout evolution [1]. As signature patterns we selected two conserved regions. The first contains a perfectly conserved tetrapeptide which is part of the GTP-binding pocket [2], the second region also contains conserved residues involved in GTP-binding. Description of pattern(s) and/or profile(s) Consensus pattern [DEN]-[LIVM](2)-x(2)-[KRNQ]-[DEN]-[LIVM]-x(3)-[ST]-x-C-E- H-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Patterns and text revised. References [1] Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H.

Pfam	Prosite	Full Name	Description
			Biochem. Biophys. Res. Commun. 212:705-711(1995). [2] Nar H., Huber R., Meining W., Schmid C., Weinkauf S., Bacher A. Structure 3:459-466(1995).
HCV_capsid		Hepatitis C virus capsid protein	Family members include nucleocapsid proteins of the HCV. This virus family comprises a nnucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein c and mRNA. Uses for these polypeptides include: immunulogical epitopes for vaccines; or as mRNA chaperone proteins to aid in processing or to prevent degradation. References describing examples of these capsid polypeptides include: Chen et al., Virology 188:102-113(1992); and Okamoto et al., J. Gen. Virol. 72:2697-2704(1991
HD		HD domain	Accession number: PF01966 Definition: HD domain Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -1 -1 Trusted cutoffs: -0.50 -0.50 Noise cutoffs: -2.50 -2.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99085258 Reference Title: The HD domain defines a new superfamily of metal-dependent Reference Title: phosphohydrolases. Reference Author: Aravind L, Koonin EV; Reference Location: Trends Biochem Sci 1998;23:469-472. Database Reference INTERPRO; IPR002819; Database reference: PFAMB; PB005654; Database reference: PFAMB; PB006725; Database reference: PFAMB; PB009617; Database reference: PFAMB; PB012663; Database reference: PFAMB; PB035384; Database reference: PFAMB; PB040597; Comment: HD domains are metal dependent phosphohydrolases. Number of members: 63
HDV_ag		Hepatitis delta virus delta antigen	Accession number: PF01517 Definition: Hepatitis delta virus delta antigen Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B 808 (release 4.0) Gathering cutoffs: -8 -8 Trusted cutoffs: 23.30 23.30 Noise cutoffs: -40.50 -40.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94065676 Reference Title: Characterization of RNA-binding domains of hepatitis delta Reference Title: antigen. Reference Author: Poisson F, Roingeard P, Baillou A, Dubois F, Bonelli F, Reference Author: Calogero RA, Goudeau A; Reference Location. J Gen Virol 1993;74:2473-2478. Reference Number: [2] Reference Medline: 98362586 Reference Title: Structural basis of the oligomerization of hepatitis delta Reference Title: antigen. Reference Author: Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM; Reference Location: Structure 1998;6:821-830. Database Reference: SCOP; 1a92; fa; [SCOP-USA][CATH-PDBSUM]

Pfam	Prosite	Full Name	Description
			Database Reference INTERPRO; IPR002506; Database Reference PDB; 1a92 A; 12; 23; Database Reference PDB; 1a92 B; 12; 23; Database Reference PDB; 1a92 C; 12; 23; Database Reference PDB; 1a92 D; 12; 60; Database Reference PDB; 1a92 A; 47; 60; Database Reference PDB; 1a92 B; 47; 60; Database Reference PDB; 1a92 C; 47; 60; Comment: The hepatitis delta virus (HDV) encodes a single protein, the Comment: hepatitis delta antigen (HDAg). The central region of this protein Comment: has been shown to bind RNA [1]. Several interactions are also Comment: mediated by a coiled-coil region at the N terminus of the protein [2]. Number of members: 145
hemolysinCabind	PDOC00293	Hemolysin-type calcium-binding region signature	Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which are currently known to belong to this category are: - Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: E. coli (gene hlyA), A. pleuropneumoniae (gene appA), A. actinomycetemcomitans and P. haemolytica (leukotoxin) (gene lktA). - Cyclolysin from Bordetella pertussis (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin. - Extracellular zinc proteases: serralysin (EC 3.4.24.40) from Serratia, prtB and prtC from Erwinia chrysanthemi and aprA from Pseudomonas aeruginosa. - Nodulation protein nodO from Rhizobium leguminosarum. We derived a signature pattern from conserved positions in the sequence of the calcium-binding domain. Description of pattern(s) and/or profile(s) Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins. Last update October 1993 / Text revised. References [1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).

Pfam	Prosite	Full Name	Description
			[2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).
Heptosyltranf		Heptosyltransferase	Accession number: PF01075 Definition: Heptosyltransferase Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_839 (release 3.0) Gathering cutoffs: -40 -40 Trusted cutoffs: -31.80 -31.80 Noise cutoffs: -47.10 -47.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98112827 Reference Title: Enzymatic synthesis of lipopolysaccharide in Escherichia Reference Title: coli. Purification and properties of heptosyltransferase I. Reference Author: Kadrmas JL, Raetz CR; Reference Location: J Biol Chem 1998;273:2799-2807. Database Reference INTERPRO; IPR002201; Database reference: PFAMB; PB021100; Database reference: PFAMB; PB033445; Database reference: PFAMB; PB041423; Comment: Lipopolysaccharide is a major component of the outer leaflet of Comment: the outer membrane in Gram-negative bacteria. It is composed of Comment: three domains; lipid A, Core oligosaccharide and the O-antigen. Comment: All of these enzymes transfer heptose to the lipopolysaccharide Comment: core. Number of members: 46
Herpes_alk_exo		Herpesvirus alkaline exonuclease	Accession number: PF01771 Definition: Herpesvirus alkaline exonuclease Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_822 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 318.00 318.00 Noise cutoffs: -277.60 -277.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 85107093 Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Banks LM, Halliburton IW, Purifoy DJ, Killington RA, Powell Reference Author: KL; Reference Location: J Gen Virol 1985;66:1-14. Database Reference INTERPRO; IPR001616; Comment: This family includes various alkaline exonucleases from Comment: members of the herpesviridae. Alkaline exonuclease Comment: appears to have an important role in the replication of Comment: herpes simplex virus [1]. Number of members: 23
Herpes_gI		Alphaherpesvirus glycoprotein I	Accession number: PF01688 Definition: Alphaherpesvirus glycoprotein I Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1222 (release 4.1) Gathering cutoffs: 25 25

Pfam	Prosite	Full Name	Description
			Trusted cutoffs: 157.20 157.20 Noise cutoffs: -126.70 -126.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96357074 Reference Title: Biosynthesis of glycoproteins E and I of feline Reference Title: herpesvirus: gE-gI interaction is required for Reference Title: intracellular transport. Reference Author: Mijnes JD, van der Horst LM, van Anken E, Horzinek MC, Reference Author: Rottier PJ, de Groot RJ; Reference Location: J Virol 1996;70:5466-5475. Reference Number: [2] Reference Medline: 94267406 Reference Title: Identification of the feline herpesvirus type 1 (FHV-1) Reference Title: genes encoding glycoproteins G, D, I and E: expression of Reference Title: FHV-1 glycoprotein D in vaccinia and raccoon poxviruses. Reference Author: Spatz SJ, Rota PA, Maes RK; Reference Location: J Gen Virol 1994;75:1235-1244. Reference Number: [3] Reference Medline: 94267879 Reference Title: Unusual phosphorylation sequence in the gpIV (gI) component Reference Title: of the varicella-zoster virus gpI-gpIV glycoprotein complex Reference Title: (VZV gE-gI complex). Reference Author: Yao Z, Grose C; Reference Location: J Virol 1994;68:4204-4211. Database Reference INTERPRO; IPR002874; Comment: This family consists of glycoprotein I form various members of the Comment: alphaherpesvirinae these include herpesvirus, varicella-zoster virus Comment: and pseudorabies virus. Glycoprotein I (gI) is important during natural Comment: infection, mutants lacking gI produce smaller lesions at the site of Comment: infection and show reduced neuronal spread [1]. gI forms a heterodimeric Comment: complex with gE; this complex displays Fc receptor activity (binds to Comment: the Fc region of immunoglobulin) [1]. Glycoproteins are also important Comment: in the production of virus-neutralizing antibodies and cell mediated Comment: immunity [2]. The alphaherpesvirinae have a dsDNA gnome and have no Comment: RNA stage during viral replication Number of members: 22
Herpes_glycop_D		Herpesvirus glycoprotein M	Accession number: PF01528 Definition: Herpesvirus glycoprotein M Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_929 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 197.30 197.30 Noise cutoffs: -229.70 -229.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96357105 Reference Title: Identification and characterization of pseudorabies virus Reference Title: glycoprotein gM as a nonessential virion component. Reference Author: Dijkstra JM, Visser N, Mettenleiter TC, Klupp BG; Reference Location: J Virol 1996;70:5684-5688.

Pfam	Prosite	Full Name	Description
			Reference Number: [2] Reference Medline: 95381611 Reference Title: Identification and molecular characterization of the murine Reference Title: cytomegalovirus homolog of the human cytomegalovirus UL100 Reference Title: gene. Reference Author: Li W, Eidman K, Gehrz RC, Kari B; Reference Location: Virus Res 1995;36:163-175. Database Reference INTERPRO; IPR000785; Comment: The herpesvirus glycoprotein M (gM) is an integral membrane protein Comment: predicted to contain 8 transmembrane segments [2]. Glycoprotein M is Comment: not essential for viral replication [1]. Number of members: 24
HesB-like	PDOC00887	Hypothetical hesB/yadR/yfhF family signature	The following uncharacterized proteins have been shown [1] to share regions of similarities: - Anabaena and related cyanobacteria protein hesB which may be required for nitrogen fixation. - Escherichia coli hypothetical protein yadR and HI1723, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein ydiC. - Escherichia coli hypothetical protein yfhF and HI0376, the corresponding Haemophilus influenzae protein. - Mycobacterium tuberculosis hypothetical protein Rv2204c. - Synechocystis strain PCC 6803 hypothetical protein slr1417. - Synechocystis strain PCC 6803 hypothetical protein slr1565. - A hypothetical protein in the nifU 5'region of many nitrogen fixing bacteria. - Porphyra purpurea chloroplast hypothetical protein in apcF-rps4 intergenic region. - Yeast hypothetical protein YLL027W. - Yeast hypothetical protein YPR067W. These are small proteins (106 to 135 amino-acid residues in bacteria, about 200 residues in fungi) that contain a number of conserved regions. The most noteworthy of these regions is located in the C-terminal extremity, it contains two conserved cysteines. We have used this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern F-x-[LIVMFY]-x-N-[PG]-[NSKQ]-x(4)-C-x-C-[GS]-x-S-F Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Bairoch A., Rudd K.E. Unpublished observations (1995).
HisG	PDOC01020	ATP phosphoribosyltransferase signature	ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we

Pfam	Prosite	Full Name	Description
			selected a region located in the C-terminal part of this enzyme. Description of pattern(s) and/or profile(s) Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / First entry.
histone	PDOC00045 PDOC00046 PDOC00287 PDOC00308	Histone H2A signature; Histone H4 signature; Histone H3 signatures; Histone H2B signature	Histone H2A is one of the four histones, along with H2B, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2A sequences [1,2,E1] we selected, as a signature pattern, a conserved region in the N-terminal part of H2A. This region is conserved both in classical S- phase regulated H2A's and in variant histone H2A's which are synthesized throughout the cell cycle. Description of pattern(s) and/or profile(s) Consensus pattern [AC]-G-L-x-F-P-V Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update November 1995 / Pattern and text revised. References [1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991). [2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994). [E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html Histone H4 is one of the four histones, along with H2A, H2B and H3, which forms the eukaryotic nucleosome core. Along with H3, it plays a central role in nucleosome formation. The sequence of histone H4 has remained almost invariant in more then 2 billion years of evolution [1,E1]. The region we use as a signature pattern is a pentapeptide found in positions 14 to 18 of all H4 sequences. It contains a lysine residue which is often acetylated [2] and a histidine residue which is implicated in DNA-binding [3]. Description of pattern(s) and/or profile(s) Consensus pattern G-A-K-R-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 1. Last update November 1995 / Text revised.

Pfam	Prosite	Full Name	Description
			References [1] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994). [2] Doenecke D., Gallwitz D. Mol. Cell. Biochem. 44:113-128(1982). [3] Ebralidse K.K., Grachev S.A., Mirzabekov A.D. Nature 331:365-367(1988). [E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html Histone H3 is one of the four histones, along with H2A, H2B and H4, which forms the eukaryotic nucleosome core. It is a highly conserved protein of 135 amino acid residues [1,2,E1]. The following proteins have been found to contain a C-terminal H3-like domain: - Mammalian centromeric protein CENP-A [3]. Could act as a core histone necessary for the assembly of centromeres. - Yeast chromatin-associated protein CSE4 [4]. - Caenorhabditis elegans chromosome III encodes two highly related proteins (F54C8.2 and F58A4.3) whose C-terminal section is evolutionary related to the last 100 residues of H3. The function of these proteins is not yet known. We developed two signature patterns, The first one corresponds to a perfectly conserved heptapeptide in the N-terminal part of H3. The second one is derived from a conserved region in the central section of H3. Description of pattern(s) and/or profile(s) Consensus pattern K-A-P-R-K-Q-L Sequences known to belong to this class detected by the pattern ALL, except for the H3-like proteins and some protozoan H3. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991). [2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994). [3] Sullivan K.F., Hechenberger M., Masri K. J. Cell Biol. 127:581-592(1994). [4] Stoler S., Keith K.C., Curnick K.E., Fitzgerald-Hayes M.

Pfam	Prosite	Full Name	Description
			Genes Dev. 9:573-586(1995). [E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html Histone H2B is one of the four histones, along with H2A, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2B sequences [1,2,E1], we selected a conserved region in the C-terminal part of H2B. Description of pattern(s) and/or profile(s) Consensus pattern [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L- x-[KR]-H-A-[LIVM]-[STA]-E-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991). [2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994). [E1] http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES/index.html
HMA	PDOC00804	Heavy-metal-associated domain	A conserved domain of about 30 amino acid residues has been found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in: - A variety of cation transport ATPases (E1-E2 ATPases) (see <PDOC00139>). The human copper ATPAses ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from Enterococcus faecalis and synA from Synechococcus contain one copy of the HMA domain. The cadmium ATPases cadA from Bacillus firmus and from plasmid pl258 from Staphylococcus aureus also contain a single HMA domain, while a chromosomal Staphylococcus aureus cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixl from Rhizobium meliloti, pacS from Synechococcus strain PCC 7942), Mycobacterium leprae ctpA and ctpB and Escherichia coli hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal section. - Mercuric reductase (EC 1.16.1.1) (gene merA) which is generally encoded by plasmids carried by mercury-resistant Gram-negative

Pfam	Prosite	Full Name	Description
			bacteria. Mercuric reductase is a class-1 pyridine nucleotide-disulphide oxidoreductase (see <PDOC00073>). There is generally one HMA domain (with the exception of a chromosomal merA from Bacillus strain RC607 which has two) in the N-terminal part of merA. - Mercuric transport protein periplasmic component (gene merP), also encoded by plasmids carried by mercury-resistant Gram-negative bacteria. It seems to be a mercury scavenger that specifically binds to one Hg(2+) ion and which passes it to the mercuric reductase via the merT protein. The N-terminal half of merP is a HMA domain. - Helicobacter pylori copper-binding protein copP. - Yeast protein ATX1 [2], which could act in the transport and/or partitioning of copper. The consensus pattern for HMA spans the complete domain. Description of pattern(s) and/or profile(s) Consensus pattern [LIVNS]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]- x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS] [The two C's probably bind metals] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 6. Last update December 1999 / Pattern and text revised. References [1] Bull P.C., Cox D.W. Trends Genet. 10:246-252(1994). [2] Lin S.-J., Culotta V.L. Proc. Natl. Acad. Sci. U.S.A. 92:3784-3788(1995).
HMG-CoA_red	PDOC00064	Hydroxymethylglutaryl-coenzyme A reductase signatures and profile	Hydroxymethylglutaryl-coenzyme A reductase (EC 1.1.1.34) (HMG-CoA reductase) [1,2] catalyzes the NADP-dependent synthesis of mevalonate from 3-hydroxy-3-methylglutaryl-CoA. In vertebrates, HMG-CoA reductase is the rate-limiting enzyme in cholesterol biosynthesis. In plants, mevalonate is the precursor of all isoprenoid compounds. HMG-CoA reductase is a membrane bound enzyme. Structurally, it consists of 3 domains. An N-terminal region that contains a variable number of transmembrane segments (7 in mammals, insects and fungi; 2 in plants), a linker region and a C-terminal catalytic domain of approximately 400 amino-acid residues. In archebacteria [3] HMG-CoA reductase, which is involved in the biosynthesis of the isoprenoids side chains of lipids, seems to be cytoplasmic and lack the N-terminal hydrophobic domain. Some bacteria, such as Pseudomonas mevalonii, can use mevalonate as the sole carbon source. These bacteria use an NAD-dependent HMG-CoA reductase

Pfam	Prosite	Full Name	Description
			(EC 1.1.1.88) to deacetylate mevalonate into 3-hydroxy-3-methylglutaryl-CoA [3]. The Pseudomonas enzyme is structurally related to the catalytic domain of NADP-dependent HMG-CoA reductases. We selected three conserved regions as signature patterns for HMG-CoA reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base. Description of pattern(s) and/or profile(s) Consensus pattern [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5. Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x- [FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaebacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update November 1997 / Patterns and text revised; profile added. References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989). [2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988). [3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992). [4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989). [5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).
HMGL-like	PDOC00813 PDOC00643	Hydroxymethylglutaryl-coenzyme A lyase active site; Alpha-isopropylmalate and homocitrate	3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In vertebrates it is a mitochondrial enyme which is involved in

Pfam	Prosite	Full Name	Description
		synthases signatures	ketogenesis and in leucine catabolism [1]. In some bacteria, such as Pseudomonas mevalonii, it is involved in mevalonate catabolism (gene mvaB). A cysteine has been shown [2], in mvaB, to be required for the activity of the enzyme. The region around this residue is perfectly conserved and is used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern S-V-A-G-L-G-G-C-P-Y [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / First entry. References [1] Mitchell G.A., Robert M.-F., Hruz P.W., Wang S., Fontaine G., Behnke C.E., Mende-Mueller L.M., Schappert K., Lee C., Gibson K.M., Miziorko H.M. J. Biol. Chem. 268:4376-4381(1993). [2] Hruz P.W., Narasimhan C., Miziorko H.M. Biochemistry 31:6842-6847(1992). The following enzymes have been shown [1] to be functionally as well as evolutionary related: - Alpha-isopropylmalate synthase (EC 4.1.3.12) which catalyzes the first step in the biosynthesis of leucine, the condensation of acetyl-CoA and alpha- ketoisovalerate to form 2-isopropylmalate synthase. - Homocitrate synthase (EC 4.1.3.21) (gene nifV) which is involved in the biosynthesis of the iron-molybdenum cofactor of nitrogenase and catalyzes the condensation of acetyl-CoA and alpha-ketoglutarate into homocitrate. - Soybean late nodulin 56. - Methanococcus jannaschii hypothetical proteins MJ0503, MJ1195 and MJ1392. We have selected two conserved regions as signature patterns for these enzymes. The first region is located in the N-terminal section while the second region is located in the central section and contains two conserved histidine residues which could be implicated in the catalytic mechanism. Description of pattern(s) and/or profile(s) Consensus pattern L-R-[DE]-G-x-Q-x(10)-K Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI] Sequences known to belong to this class detected by the pattern ALL.

Pfam	Prosite	Full Name	Description
			Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Wang S.-Z., Dean D.R., Chen J.-S., Johnson J.L. J. Bacteriol. 173:3041-3046(1991).
hormone5	PDOC00237	Neurohypophysial hormones signature	Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels. Like the majority of active peptides, both hormones are synthesized as larger protein precursors that are enzymatically converted to their mature forms. Peptides belonging to this family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin, glumitocin, aspargtocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin), octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs (conopressins G and S) [2]. The pattern developed to detect this category of peptides spans their entire sequence and includes four invariant amino acid residues. Description of pattern(s) and/or profile(s) Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide bond]. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Pattern and text revised. References [1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988). [2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein Res. 45:482-487(1995).
HPPK	PDOC00631	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature	All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents such as trimethoprim or sulfonamides. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate.

Pfam	Prosite	Full Name	Description
			This is the first step in a three-step pathway leading to 7,8-dihydrofolate. Bacterial HPPK (gene folK or sulD) [1] is a protein of 160 to 270 amino acids. In the lower eukaryote Pneumocystis carinii, HPPK is the central domain of a multifunctional folate synthesis enzyme (gene fas) [2]. As a signature for HPPK, we selected a conserved region located in the central section of these enzymes. Description of pattern(s) and/or profile(s) Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992). [2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).
HTH_AraC	PDOC00040	Bacterial regulatory proteins, araC family signature and profile	The many bacterial transcription regulation proteins which bind DNA through a 'helix-turn-helix' motif can be classified into subfamilies on the basis of sequence similarities. One of these subfamilies groups together the following proteins [1,2]: - aarP, a transcriptional activator of the 2'-N-acetyltransferase gene in Providencia stuartii. - ada, an Escherichia coli and Salmonella typhimurium bifunctional protein that repairs alkylated guanine in DNA by transferring the alkyl group at the O(6) position to a cysteine residue in the enzyme. The methylated protein acts a positive regulator of its own synthesis and of the alkA, alkB and aidB genes. - adaA, a Bacillus subtilis bifunctional protein that acts both as a transcriptional activator of the ada operon and as a methylphosphotriester-DNA alkyltransferase. - adiY, an Escherichia coli protein of unknown function. - aggR, the transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative Escherichia coli. - appY, a protein which acts as a transcriptional activator of acid phosphatase and other proteins during the deceleration phase of growth and acts as a repressor for other proteins that are synthesized in exponential growth or in the stationary phase. - araC, the arabinose operon regulatory protein, which activates the transcription of the araBAD genes. - cafR, the Yersinia pestis F1 operon positive regulatory protein.

Pfam	Prosite	Full Name	Description
			- celD, the Escherichia coli cel operon repressor. - cfaD, a protein which is required for the expression of the CFA/I adhesin of enterotoxigenic Escherichia coli. - csvR, a transcriptional activator of fimbrial genes in enterotoxigenic Escherichia coli. - envY, the porin thermoregulatory protein, which is involved in the control of the temperature-dependent expression of several Escherichia coli envelope proteins such as ompF, ompC, and lamB. - exsA, an activator of exoenzyme S synthesis in Pseudomonas aeruginosa. - fapR, the positive activator for the expression of the 987P operon coding for the fimbrial protein in enterotoxigenic Escherichia coli. - hrpB, a positive regulator of pathogenicity genes in Burkholderia solanacearum. - invF, the Salmonella typhimurium invasion operon regulator. - marA, which may be a transcriptional activator of genes involved in the multiple antibiotic resistance (mar) phenotype. - melR, the melibiose operon regulatory protein, which activates the transcription of the melAB genes. - mixE, a Shigella flexneri protein necessary for secretion of ipa invasins. - mmsR, the transcriptional activator for the mmsAB operon in Pseudomonas aeruginosa. - msmR, the multiple sugar metabolism operon transcriptional activator in Streptococcus mutans. - pchR, a Pseudomonas aeruginosa activator for pyochelin and ferripyochelin receptor. - perA, a transcriptional activator of the eaeA gene for intimin in enteropathogenic Escherichia coli. - pocR, a Salmonella typhimurium regulator of the cobalamin biosynthesis operon. - pqrA, from Proteus vulgaris. - rafR, the regulator of the raffinose operon in Pediococcus pentosaceus. - ramA, from Klebsiella pneumoniae. - rhaR, the Escherichia coli and Salmonella typhimurium L-rhamnose operon transcriptional activator. - rhaS, an Escherichia coli and Salmonella typhimurium positive activator of genes required for rhamnose utilization. - rns, a protein which is required for the expression of the cs1 and cs2 adhesins of enterotoxigenic Escherichia coli. - rob, a protein which binds to the right arm of the replication origin oriC of the Escherichia coli chromosome. - soxS, a protein that, with the soxR protein, controls a superoxide response regulon in Escherichia coli. - tetD, a protein from transposon TN10. - tcpN or toxT, the Vibrio cholerae transcriptional activator of the tcp operon involved in pilus biosynthesis and transport. - thcR, a probable regulator of the thc operon for the degradation of the thiocarbamate herbicide EPTC in Rhodococcus sp. strain NI86/21. - ureR, the transcriptional activator of the plasmid-encoded urease operon in Enterobacteriaceae.

Pfam	Prosite	Full Name	Description
			- virF and lcrF, the Yersinia virulence regulon transcriptional activator. - virF, the Shigella transcriptional factor of invasion related antigens ipaBCD. - xylR, the Escherichia coli xylose operon regulator. - xylS, the transcriptional activator of the Pseudomonas putida TOL plasmid (pWWO, pWW53 and pDK1) meta operon (xylDLEGF genes). - yfeG, an Escherichia coli hypothetical protein. - yhiW, an Escherichia coli hypothetical protein. - yhiX, an Escherichia coli hypothetical protein. - yidL, an Escherichia coli hypothetical protein. - yijO, an Escherichia coli hypothetical protein. - yuxC, a Bacillus subtilis hypothetical protein. - yzbC, a Bacillus subtilis hypothetical protein. Except for celD, all of these proteins seem to be positive transcriptional factors. Their size range from 107 (soxS) to 529 (yzbC) residues. The helix-turn-helix motif is located in the third quarter of most of the sequences; the N-terminal and central regions of these proteins are presumed to interact with effector molecules and may be involved in dimerization [3]. The minimal DNA binding domain, which spans roughly 100 residues and comprises the HTH motif contains another region with similarity to classical HTH domain. However, it contains an insertion of one residue in the turn-region. A signature pattern was derived from the region that follows the first HTH domain and that includes the totality of the putative second HTH domain. A more sensitive detection of members of the araC family is available through the use of a profile which spans the minimal DNA-binding region of 100 residues. Description of pattern(s) and/or profile(s) Consensus pattern [KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]- x(4,9)-[LIVMF]-x(2)-[LIVMSTA]-[GSTACIL]-x(3)-[GANQRF]- [LIVMFY]-x(4,5)-[LFY]-x(3)-[FYIVA]-{FYWHCM}-x(3)-[GSADENQKR]-x-[NSTAPKL]-[PARL] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 37. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Ramos J.L. jlramos@samba.cnb.uam.es Gallegos M.-T. mtrini@samba.cnb.uam.es Last update November 1997 / Text revised. References [1] Gallegos M.-T., Michan C., Ramos J.L.

Pfam	Prosite	Full Name	Description
			Nucleic Acids Res. 21:807-810(1993). [2] Henikoff S., Wallace J.C., Brown J.P. Meth. Enzymol. 183:111-132(1990). [3] Bustos S.A., Schleif R.F. Proc. Natl. Acad. Sci. U.S.A. 90:5638-5642(1993).
Hydrolase		haloacid dehalogenase-like hydrolase	Accession number: PF00702 Definition: haloacid dehalogenase-like hydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_566 (release 2.1) Gathering cutoffs: 7 7 Trusted cutoffs: 7.10 7.10 Noise cutoffs: 2.90 2.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96355356 Reference Title: Crystal structure of L-2-haloacid dehalogenase from Reference Title: Pseudomonas sp. YL. An alpha/beta hydrolase structure that Reference Title: is different from the alpha/beta hydrolase fold. Reference Author: Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara T, Esaki N, Reference Author: Soda K; Reference Location: J Biol Chem 1996;271:20322-20330. Database Reference: SCOP; 1jud; sf; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001454; Database Reference PDB; 1jud ; 4; 197; Database Reference PDB; 1zrm ; 4; 197; Database Reference PDB; 1zrn ; 4; 197; Database Reference PDB; 1aq6 A; 2; 193; Database Reference PDB; 1aq6 B; 2; 193; Database Reference PDB; 1qq5 A; 2; 193; Database Reference PDB; 1qq5 B; 2; 193; Database Reference PDB; 1qq6 A; 2; 193; Database Reference PDB; 1qq6 B; 2; 193; Database Reference PDB; 1qq7 A; 2; 193; Database Reference PDB; 1qq7 B; 2; 193; Database Reference PDB; 1cqz A; 4; 19; Database Reference PDB; 1cr6 A; 4; 19; Database Reference PDB; 1cqz B; 4; 206; Database Reference PDB; 1cr6 B; 4; 206; Database Reference PDB; 1cqz A; 48; 206; Database Reference PDB; 1cr6 A; 48, 206; Database reference: PFAMB; PB000701; Database reference: PFAMB; PB001048; Database reference: PFAMB; PB019234; Database reference: PFAMB; PB032787; Database reference: PFAMB; PB040985; Database reference: PFAMB; PB041061; Database reference: PFAMB; PB041182; Database reference: PFAMB; PB041477; Database reference: PFAMB; PB041535; Database reference: PFAMB; PB041628; Database reference: PFAMB; PB041677; Comment: This family are structurally different from the alpha/ Comment: beta hydrolase family (abhydrolase). Comment: This family includes L-2-haloacid dehalogenase, epoxide Comment: hydrolases and phosphatases. Comment: The structure of the family consists of two domains. One Comment: is an inserted four helix bundle, which is the least well

Pfam	Prosite	Full Name	Description
			Comment: conserved region of the alignment, between residues 16 and Comment: 96 of Swiss:P24069. The rest of the fold is composed of the Comment: core alpha/beta domain. Number of members: 134
HypB_UreG		HypB/UreG nucleotide-binding domain	Accession number: PF01495 Definition: HypB/UreG nucleotide-binding domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_428 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 197.70 197.70 Noise cutoffs: -40.00 -40.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97285753 Reference Title: The HypB protein from Bradyrhizobium japonicum can store Reference Title: nickel and is required for the nickel-dependent Reference Title: transcriptional regulation of hydrogenase. Reference Author: Olson JW, Fu C, Maier RJ; Reference Location: Mol Microbiol 1997;24:119-128. Reference Number: [2] Reference Medline: 97352660 Reference Title: Characterization of UreG, identification of a Reference Title: UreD-UreF-UreG complex, and evidencesuggesting that a Reference Title: nucleotide-binding site in UreG is required for in vivo Reference Title: metallocenter assembly of Klebsiella aerogenes urease. Reference Author: Moncrief MB, Hausinger RP; Reference Location: J Bacteriol 1997;179:4081-4086. Reference Number: [3] Reference Medline: 93139028 Reference Title: The product of the hypB gene, which is required for nickel Reference Title: incorporation into hydrogenases, is a novel guanine Reference Title: nucleotide-binding protein. Reference Author: Maier T, Jacobi A, Sauter M, Bock A; Reference Location: J Bacteriol 1993;175:630-635. Reference Number: [4] Reference Medline: 92325016 Reference Title: Klebsiella aerogenes urease gene cluster: sequence of ureD Reference Title: and demonstration that four accessory genes (ureD, ureE, Reference Title: ureF, and ureG) are involved in nickel metallocenter Reference Title: biosynthesis. Reference Author: Lee MH, Mulrooney SB, Renner MJ, Markowicz Y, Hausinger RP; Reference Location: J Bacteriol 1992;174:4324-4330. Database Reference INTERPRO; IPR002894, Comment: This domain is found in HypB, a hydrogenase expression / formation Comment: protein, and UreG a urease accessory protein. Both these proteins contain Comment: a P-loop nucleotide binding motif [2,3]. HypB has GTPase activity Comment: and is a guanine nucleotide binding protein [3]. It is not known Comment: whether UreG binds GTP or some other nucleotide. Both enzymes are involved Comment: in nickel binding. HypB can store nickel and is required for nickel Comment: dependent hydrogenase expression [1]. UreG is required for functional Comment: incorporation of the urease nickel

Pfam	Prosite	Full Name	Description
			metallocenter.[4] GTP hydrolysis may Comment: required by these proteins for nickel incorporation into other nickel Comment: proteins [1]. Number of members: 41
IBB		Importin beta binding domain	Accession number: PF01749 Definition: Importin beta binding domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_544 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 67.30 67.30 Noise cutoffs: -15.90 -15.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98359119 Reference Title: Crystallographic analysis of the recognition of a nuclear Reference Title: localization signal by the nuclear import factor Reference Title: karyopherin alpha. Reference Author: Conti E, Uy M, Leighton L, Blobel G, Kuriyan J; Reference Location: Cell 1998;94:193-204. Reference Number: [2] Reference Medline: 98275030 Reference Title: Importins and exportins: how to get in and out of the Reference Title: nucleus [published erratum appears in Trends Biochem Sci Reference Title: 1998 Jul;23(7):235] Reference Author: Weis K; Reference Location: Trends Biochem Sci 1998;23:185-189. Reference Number: [3] Reference Medline: 98250643 Reference Title: Transport into and out of the cell nucleus. Reference Author: Gorlich D; Reference Location: EMBO J 1998;17:2721-2727. Reference Number: [4] Reference Medline: 96270582 Reference Title: The binding site of karyopherin alpha for karyopherin beta Reference Title: overlaps with a nuclear localization sequence. Reference Author: Moroianu J, Blobel G, Radu A; Reference Location: Proc Natl Acad Sci U S A 1996;93:6572-6576. Reference Number: [5] Reference Medline: 96203101 Reference Title: A 41 amino acid motif in importin-alpha confers binding to Reference Title: importin- beta and hence transit into the nucleus. Reference Author: Gorlich D, Henklein P, Laskey RA, Hartmann E; Reference Location: EMBO J 1996;15:1810-1817. Database Reference: SCOP; 1bk5; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002652; Database Reference PDB; 1ejl I; 72; 99; Database Reference PDB; 1ejy I; 72; 99; Database Reference PDB; 1ial A; 44; 99; Database Reference PDB; 1qgr B; 28; 51; Database Reference PDB; 1qgk B; 11; 54; Database Reference PDB; 1ee5 A; 90; 110; Database Reference PDB; 1bk5 A; 89; 110; Database Reference PDB; 1bk5 B; 89; 110; Database Reference PDB; 1bk6 A; 89; 110; Database Reference PDB; 1bk6 B; 89; 110; Database Reference PDB; 1ee4 A; 87; 110; Database Reference PDB; 1ee4 B; 87; 110; Comment: This family consists of the importin alpha

Pfam	Prosite	Full Name	Description
			(karyopherin alpha), Comment: importin beta (karyopherin beta) binding domain. The domain mediates Comment: formation of the importin alpha beta complex; required for classical Comment: NLS import of proteins into the nucleus, through the nuclear pore Comment: complex and across the nuclear envelope. Comment: Also in the alignment is the NLS of importin alpha which overlaps Comment: with the IBB domain [4]. Number of members: 38
IF-2B		Initiation factor 2 subunit family	Accession number: PF01008 Definition: Initiation factor 2 subunit family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1302 (release 3.0) Gathering cutoffs: -135 -135 Trusted cutoffs: -82.40 -82.40 Noise cutoffs: -157.30 -157.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98188271 Reference Title: Archaeal translation initiation revisited: the initiation Reference Title: factor 2 and eukaryotic initiation factor 2B Reference Title: alpha-beta-delta subunit families. Reference Author: Kyrpides NC, Woese CR; Reference Location: Proc Natl Acad Sci U S A 1998;95:3726-3730. Database Reference INTERPRO; IPR000649; Comment: This family includes initiation factor 2B alpha, beta and delta Comment: subunits from eukaryotes, initiation factor 2B subunits 1 and 2 Comment: from archaebacteria and some proteins of unknown function from Comment: prokaryotes. Initiation factor 2 binds to Met-tRNA, GTP and the Comment: small ribosomal subunit. Number of members: 33
IF3	PDOC00723	Initiation factor 3 signature	Initiation factor 3 (IF-3) (gene infC) [1] is one of the three factors required for the initiation of protein biosynthesis in bacteria. IF-3 is thought to function as a fidelity factor during the assembly of the ternary initiation complex which consist of the 30S ribosomal subunit, the initiator tRNA and the messenger RNA. IF-3 binds to the 30S ribosomal subunit; it is a basic protein of 141 to 212 residues. The chloroplast initiation factor IF-3(chl) is a protein that enhances the poly(A,U,G)-dependent binding of the initiator tRNA to chloroplast ribosomal 30s subunits. In its mature form it is a protein of about 400 residues whose central section is evolutionary related to the sequence of bacterial IF-3 [2]. As a signature pattern we selected a highly conserved region located in the central section of bacterial IF-3 and of IF-3(chl). Description of pattern(s) and/or profile(s) Consensus pattern [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-

Pfam	Prosite	Full Name	Description
			[LIVMFYS]-x-[FY]- [DEQTH]-x(2)-[KRQ] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Liveris D., Schwartz J.J., Geertman R., Schwartz I. FEMS Microbiol. Lett. 112:211-216(1993). [2] Lin Q., Ma L., Burkhart W., Spremulli L.L. J. Biol. Chem. 269:9436-9444(1994).
IF4E	PDOC00641	Eukaryotic initiation factor 4E signature	Eukaryotic translation initiation factor 4E (eIF-4E) [1] is a protein that binds to the cap structure of eukaryotic cellular mRNAs. eIF-4E recognizes and binds the 7-methylguanosine-containing (m7Gppp) cap during an early step in the initiation of protein synthesis and facilitates ribosome binding to a mRNA by inducing the unwinding of its secondary structures. eIF-4E is a conserved protein of about 25 Kd. Site directed mutagenesis experiments have shown [2] that a tryptophan in the central part of the sequence of human eIF-4E seems to be implicated in cap-binding. The signature pattern for eIF-4E includes this tryptophan. Description of pattern(s) and/or profile(s) Consensus pattern [DE]-[IFY]-x(2)-F-[KR]-x(2)-[LIVM]-x-P-x-W-E-[DVA]-x(5)-G- G-[KR]-W [The first W seems to be involved in cap-binding] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Thach R.E. Cell 68:177-180(1992). [2] Ueda H., Iyo H., Doi M., Inoue M., Ishida T., Morioka H., Tanaka T., Nishikawa S., Uesugi S. FEBS Lett. 280:207-210(1991).
IF5_eIF4_eIF2		eIF4-gamma/eIF5/eIF2-epsilon	Accession number: PF02020 Definition: eIF4-gamma/eIF5/eIF2-epsilon Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 26.10 26.10 Noise cutoffs: -21.50 -21.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96060092 Reference Title: Multidomain organization of eukaryotic guanine nucleotide Reference Title: exchange translation initiation factor eIF-2B subunits Reference Title: revealed by analysis of conserved sequence motifs. Reference Author: Koonin EV;

Pfam	Prosite	Full Name	Description
			Reference Location: Protein Sci 1995;4:1608-1617. Comment: This domain of unknown function is found at the C-terminus Comment: of several transcription initiation factors [1]. Number of members: 31
ig	PDOC00262	Immunoglobulins and major histocompatibility complex proteins signature	The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4). The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail. It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. We developed a small pattern around the C-terminal cysteine involved in this disulfide bond which can be used to detect these category of Ig related proteins. Description of pattern(s) and/or profile(s) Consensus pattern [FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region : All, in CH2 and CH3. Ig heavy chains type Delta C region : All, in CH3. Ig heavy chains type Epsilon C region: All, in CH1, CH3 and CH4. Ig heavy chains type Gamma C region : All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region : All, in CH2, CH3 and CH4. Ig light chains type Kappa C region : In all CL except rabbit and Xenopus. Ig light chains type Lambda C region : In all CL except rabbit. MHC class I alpha chains : All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin : All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains. Other sequence(s) detected in SWISS-PROT 71. Last update May 1991 / Text revised. References [1] Gough N. Trends Biochem. Sci. 6:203-205(1981). [2] Klein J., Figueroa F. Immunol. Today 7:41-44(1986). [3]

Pfam	Prosite	Full Name	Description
			Figueroa F., Klein J. Immunol. Today 7:78-81(1986). [4] Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L. Nature 282:266-270(1979). [5] Cushley W., Owen M.J. Immunol. Today 4:88-92(1983). [6] Beck S., Barrel B.G. Nature 331:269-272(1988).
IMPDH_C	PDOC00391	IMP dehydrogenase / GMP reductase signature	IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2]. GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides. IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. We have used this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988). [2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990). [3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).
Inos-1-P_synth		Myo-inositol-1-phosphate synthase	Accession number: PF01658 Definition: Myo-inositol-1-phosphate synthase Author: Bashton M, Bateman A Alignment method of seed: Clustalw

Pfam	Prosite	Full Name	Description
			Source of seed members: Pfam-B_959 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 86.80 86.80 Noise cutoffs: -219.00 -219.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95066381 Reference Title: Comparison of INO1 gene sequences and products in Candida Reference Title: albicans and Saccharomyces cerevisiae. Reference Author: Klig LS, Zobel PA, Devry CG, Losberger C; Reference Location: Yeast 1994;10:789-800. Database Reference INTERPRO; IPR002587; Comment: This is a family of myo-inositol-1-phosphate synthases. Comment: Inositol-1-phosphate catalyses the conversion of glucose-6- Comment: phosphate to inositol-1-phosphate, which is then dephosphorylated Comment: to inositol [1]. Inositol phosphates play an important role in Comment: signal transduction. Number of members: 27
IPP_isomerase		Isopentenyl-diphosphate delta-isomerase	Accession number: PF01772 Definition: Isopentenyl-diphosphate delta-isomerase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1099 (release 4.2) Gathering cutoffs: -88 -88 Trusted cutoffs: -66.70 -66.70 Noise cutoffs: -106.90 -106.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98409684 Reference Title: Differential expression of two isopentenyl pyrophosphate Reference Title: isomerases and enhanced carotenoid accumulation in a Reference Title: unicellular chlorophyte Reference Author: Sun Z, Cunningham FX Jr, Gantt E; Reference Location: Proc Natl Acad Sci U S A 1998;95:11482-11488. Reference Number: [2] Reference Medline: 97373600 Reference Title: Cloning and subcellular localization of hamster and rat Reference Title: isopentenyl diphosphate dimethylallyl diphosphate Reference Title: isomerase. A PTS1 motif targets the enzyme to peroxisomes. Reference Author: Paton VG, Shackelford JE, Krisans SK; Reference Location: J Biol Chem 1997;272:18945-18950. Database Reference INTERPRO; IPR002667; Comment: Isopentenyl-diphosphate delta-isomerase or IPP isomerase EC:5.3.3.2 Comment: catalyses the interconversion of isopentenyl diphosphate and Comment: dimethylallyl diphosphate. Dimethylallyl phosphate is the initial substrate Comment: for the biosynthesis of carotenoids and other long chain isoprenoids [1]. Number of members: 24
K-box	PDOC00302	MADS-box domain signature and profile	A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below: - Serum response factor (SRF) [1], a mammalian transcription factor that

Pfam	Prosite	Full Name	Description
			binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the transcription initiation site of genes such as c-fos. - Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. - Drosophila myocyte-specific enhancer factor 2 (MEF2). - Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes. - Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80). - Yeast transcription factor RLM1. - Yeast transcription factor SMP1. - Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower. - Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and Pistillata (PI) which act locally to specify the identity of the floral meristem and to determine sepal and petal development [4]. - Antirrhinum majus and tobacco homeotic protein deficiens (DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO cause the transformation of petals into sepals and of stamina into carpels. - Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6]. - Antirrhinum majus morphogenetic protein DEF H33 (squamosa). In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. We have derived a pattern that spans the complete length of the domain. The profile also spans the length of the MADS-box. Description of pattern(s) and/or profile(s) Consensus pattern R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)- x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x- [LIVM](3)-x(6)-[LIVMF]-x(2)-[FY] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so. Last update July 1999 / Pattern and text revised. References [1] Norman C., Runswick M., Pollock R., Treisman R.

Pfam	Prosite	Full Name	Description
			Cell 55:989-1003(1988). [2] Passmore S., Maine G.T., Elble R., Christ C., Tye B.-K. J. Mol. Biol. 204:593-606(1988). [3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. Nature 346:35-39(1990). [4] Goto K., Meyerowitz E.M. Genes Dev. 8:1548-1560(1994). [5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W.-E., Saedler H., Sommer H., Schwartz-Sommer Z. EMBO J. 11:4693-4704(1992). [6] Ma H., Yanofsky M.F., Meyerowitz E.M. Genes Dev. 5:484-495(1991). [E1] http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014
Keratin_B2		Keratin, high sulfur B2 protein	Accession number: PF01500 Definition: Keratin, high sulfur B2 protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_706 (release 4.0) Gathering cutoffs: -17 -17 Trusted cutoffs: -1.50 -1.50 Noise cutoffs: -46.00 18.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98201605 Reference Title: Structure and hair follicle-specific expression of genes Reference Title: encoding the rat high sulfur protein B2 family. Reference Author: Mitsui S, Ohuchi A, Adachi-Yamada T, Hotta M, Tsuboi R, Reference Author: Ogawa H; Reference Location: Gene 1998;208:123-129. Database Reference INTERPRO; IPR002494; Comment: High sulfur proteins are cysteine-rich proteins synthesized Comment: during the differentiation of hair matrix cells, and form hair Comment: fibers in association with hair keratin intermediate filaments [1]. Comment: This family has been divided up into four regions, with the second Comment: region containing 8 copies of a short repeat [1]. This family is Comment: also known as B2 or KAP1. Number of members: 17
ketoacyl-synt	PDOC00529	Beta-ketoacyl synthases active site	Beta-ketoacyl-ACP synthase (EC 2.3.1.41) (KAS) [1] is the enzyme that catalyzes the condensation of malonyl-ACP with the growing fatty acid chain. It is found as a component of the following enzymatic systems: - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to different enzymatic activities; beta-ketoacyl synthase is one

Pfam	Prosite	Full Name	Description
			of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the beta-ketoacyl synthase domain is located in the C-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional chain; the beta-ketoacyl synthase domain is located in the N-terminal section [2]. - The multifunctional 6-methysalicylic acid synthase (MSAS) from Penicillium patulum [3]. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which has a KAS domain in its N-terminal section. - Polyketide antibiotic synthase enzyme systems. Polyketides are secondary metabolites produced by microorganisms and plants from simple fatty acids. KAS is one of the components involved in the biosynthesis of the Streptomyces polyketide antibiotics granatacin [4], tetracenomycin C [5] and erythromycin. - Emericella nidulans multifunctional protein Wa. Wa is involved in the biosynthesis of conidial green pigment. Wa is protein of 216 Kd that contains a KAS domain. - Rhizobium nodulation protein nodE, which probably acts as a beta-ketoacyl synthase in the synthesis of the nodulation Nod factor fatty acyl chain. - Yeast mitochondrial protein CEM1. The condensation reaction is a two step process: the acyl component of an activated acyl primer is transferred to a cysteine residue of the enzyme and is then condensed with an activated malonyl donor with the concomitant release of carbon dioxide. The sequence around the active site cysteine is well conserved and can be used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for bacterial and plant beta-ketoacyl synthase III (KAS III). Other sequence(s) detected in SWISS-PROT 10. Last update November 1997 / Text revised. References [1] Kauppinen S., Siggaard-Andersen M., von Wettstein-Knowles P. Carlsberg Res. Commun. 53:357-370(1988). [2] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991). [3] Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E. Eur. J. Biochem. 192:487-498(1990). [4] Bibb M.J., Biro S., Motamedi H., Collins J.F., Hutchinson C.R. EMBO J. 8:2727-2736(1989).

Pfam	Prosite	Full Name	Description
			[5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).
KRAB		KRAB box	Accession number: PF01352 Definition: KRAB box Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 0 0 Trusted cutoffs: 1.10 1.10 Noise cutoffs: -5.40 -5.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91319563 Reference Title: Conserved KRAB protein domain identified upstream from the Reference Title: zinc finger region of Kox 8. Reference Author: Thiesen HJ, Bellefroid E, Revelant O, Martial JA; Reference Location: Nucleic Acids Res 1991;19:3996-3996. Reference Number: [2] Reference Medline: 97140325 Reference Title: A novel member of the RING finger family, KRIP-1, Reference Title: associates with the KRAB-A transcriptional repressor domain Reference Title: of zinc finger proteins. Reference Author: Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre Reference Author: JV; Reference Location: Proc Natl Acad Sci U S A 1996;93:15299-15304. Reference Number: [3] Reference Medline: 96365472 Reference Title: KAP-1, a novel corepressor for the highly conserved KRAB Reference Title: repression domain. Reference Author: Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang Reference Author: XP, Neilson EG, Rauscher FJ; Reference Location: Genes Dev 1996;10:2067-2078. Database Reference INTERPRO; IPR001909; Database reference: PFAMB; PB036541; Comment: The KRAB domain (or Kruppel-associated box) is present in Comment: about a third of zinc finger proteins containing C2H2 fingers. Comment: The KRAB domain is found to be involved in protein-protein Comment: interactions [2,3]. Comment: The KRAB domain is generally encoded by two exons. The Comment: regions coded by the two exons are known as KRAB-A and Comment: KRAB-B. Number of members: 105
lectin_legB	PDOC00278	Legume lectins signatures	Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals). Legume lectins are synthesized as precursor proteins of about

Pfam	Prosite	Full Name	Description
			230 to 260 amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal). The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain. We have developed two signature patterns specific to legume lectins: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 21. Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Last update July 1999 / Patterns and text revised. References [1] Sharon N., Lis H. FASEB J. 4:3198-320(1990). [2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).
ligase-CoA		CoA-ligases	Accession number: PF00549 Definition: CoA-ligases Author: Bateman A Alignment method of seed: Clustalw Source of seed members: SCOP Gathering cutoffs: 25 25 Trusted cutoffs: 28.70 28.70 Noise cutoffs: 14.70 14.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94193797 Reference Title: The crystal structure of succinyl-CoA synthetase from Reference Title: Escherichia coli at 2.5-A resolution. Reference Author: Wolodko WT, Fraser ME, James MN, Bridger WA; Reference Location: J Biol Chem 1994;269:10883-10890. Database Reference: SCOP; 1scu; sf; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000303; Database Reference PDB; 1cqi A; 132; 279; Database Reference PDB; 1cqi D; 132; 279; Database Reference PDB; 1cqj A; 132; 279; Database Reference PDB; 1cqj D; 132; 279; Database Reference PDB; 2scu A; 132; 279; Database Reference PDB; 2scu D; 132; 279;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1scu A; 132; 279; Database Reference PDB; 1scu D; 132; 279; Database Reference PDB; 1cqi B; 246; 385; Database Reference PDB; 1cqi E; 246; 385; Database Reference PDB; 1cqj B; 246; 385; Database Reference PDB; 1cqj E; 246; 385; Database Reference PDB; 2scu B; 246; 385; Database Reference PDB; 2scu E; 246; 385; Database Reference PDB; 1scu B; 246; 388; Database Reference PDB; 1scu E; 246; 388; Database reference: PFAMB; PB039724; Database reference: PFAMB; PB041236; Comment: -!- This family includes the CoA ligases Succinyl-CoA synthetase alpha Comment: and beta chains, malate CoA ligase and ATP-citrate lyase. Comment: Some members of the family utilise ATP others use GTP. Number of members: 76
LIM_bind		LIM-domain binding protein	Accession number: PF01803 Definition: LIM-domain binding protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1352 (release 4.2) Gathering cutoffs: -92 -92 Trusted cutoffs: 13.40 13.40 Noise cutoffs: -197.90 -197.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97477378 Reference Title: Chip, a widely expressed chromosomal protein required for Reference Title: segmentation and activity of a remote wing margin enhancer Reference Title: in Drosophila. Reference Author: Morcillo P, Rosen C, Baylies MK, Dorsett D; Reference Location: Genes Dev 1997;11:2729-2740. Reference Number: [2] Reference Medline: 97336071 Reference Title: A family of LIM domain-associated cofactors confer Reference Title: transcriptional synergism between LIM and Otx homeodomain Reference Title: proteins. Reference Author: Bach I, Carriere C, Ostendorff HP, Andersen B, Rosenfeld Reference Author: MG; Reference Location: Genes Dev 1997;11:1370-1380. Reference Number: [3] Reference Medline: 97078753 Reference Title: Interactions of the LIM-domain-binding factor Ldb1 with LIM Reference Title: homeodomain proteins. Reference Author: Agulnick AD, Taira M, Breen JJ, Tanaka T, Dawid IB, Reference Author: Westphal H; Reference Location: Nature 1996;384:270-272. Reference Number: [4] Reference Medline: 97030257 Reference Title: Nuclear LIM interactor, a rhombotin and LIM homeodomain Reference Title: interacting protein, is expressed early in neuronal Reference Title: development. Reference Author: Jurata LW, Kenny DA, Gill GN; Reference Location: Proc Natl Acad Sci U S A 1996;93:11693-11698. Database Reference INTERPRO; IPR002691; Comment: The LIM-domain binding protein, binds to the LIM domain LIM of Comment: LIM homeodomain proteins which are

Pfam	Prosite	Full Name	Description
			transcriptional regulators of Comment: development. Comment: Nuclear LIM interactor (NLI) / LIM domain-binding protein 1 (LDB1) Comment: Swiss:P70662 is located in the nuclei of neuronal cells during Comment: development, it is co-expressed with Isl1 in early motor neuron Comment: differentiation and has a suggested role in the Isl1 dependent Comment: development of motor neurons [4]. Comment: It is suggested that these proteins act synergistically to enhance Comment: transcriptional efficiency by acting as co-factors for LIM homeodomain Comment: and Otx class transcription factors both of which have essential roles Comment: in development [2]. Comment: The Drosophila protein Chip Swiss:O18353 is required for segmentation Comment: and activity of a remote wing margin enhancer [1]. Chip is a ubiquitous Comment: chromosomal factor required for normal expression of diverse genes at Comment: many stages of development [1]. It is suggested that Chip cooperates Comment: with different LIM domain proteins and other factors to structurally Comment: support remote enhancer-promoter interactions [1]. Number of members: 19
Lipase_3	PDOC00110	Lipases, serine active site	Triglyceride lipases (EC 3.1.1.3) [1] are lipolytic enzymes that hydrolyzes the ester bond of triglycerides. Lipases are widely distributed in animals, plants and prokaryotes. In higher vertebrates there are at least three tissue- specific isozymes: pancreatic, hepatic, and gastric/lingual. These three types of lipases are closely related to each other as well as to lipoprotein lipase (EC 3.1.1.34) [2], which hydrolyzes triglycerides of chylomicrons and very low density lipoproteins (VLDL). The most conserved region in all these proteins is centered around a serine residue which has been shown [3] to participate, with an histidine and an aspartic acid residue, to a charge relay system. Such a region is also present in lipases of prokaryotic origin and in lecithin-cholesterol acyltransferase (EC 2.3.1.43) (LCAT) [4], which catalyzes fatty acid transfer between phosphatidylcholine and cholesterol. We have built a pattern from that region. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-S-x-G-[GSTAC] [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 35. Note Drosophila vitellogenins are also related to lipases [5], but they have lost their active site serine. Last update November 1997 / Pattern and text revised. References

Pfam	Prosite	Full Name	Description
			[1] Chapus C., Rovery M., Sarda L., Verger R. Biochimie 70:1223-1234(1988). [2] Persson B., Bengtsson-Olivecrona G., Enerback S., Olivecrona T., Joernvall H. Eur. J. Biochem. 179:39-45(1989). [3] Blow D. Nature 343:694-695(1990). [4] McLean J., Fielding C., Drayna D., Dieplinger H., Baer B., Kohr W., Henzel W., Lawn R. Proc. Natl. Acad. Sci. U.S.A. 83:2335-2339(1986). [5] Baker M.E. Biochem. J. 255:1057-1060(1988).
Lipase_GDSL	PDOC00842	Lipolytic enzymes "G-D-S-L" family, serine active site	Recently [1], a family of lipolytic enzymes has been characterized. This family currently consist of the following proteins: - Aeromonas hydrophila lipase/phosphatidylcholine-sterol acyltransferase. - Xenorhabdus luminescens lipase 1. - Vibrio mimicus arylesterase. - Escherichia coli acyl-coA thioesterase I (gene tesA). - Vibrio parahaemolyticus thermolabile hemolysin/atypical phospholipase. - Rabbit phospholipase AdRab-B, an intestinal brush border protein with esterase and phospholipase A/lysophospholipase activity that could be involved in the uptake of dietary lipids. AdRab-B contains four repeats of about 320 amino acids. - Arabidopsis thaliana and Brassic napus anther-specific proline-rich protein APG. - A Pseudomonas putida hypothetical protein in trpE-trpG intergenic region. A serine has been identified a part of the active site in the Aeromonas, Vibrio mimicus and Escherichia coli enzymes. It is located in a conserved sequence motif that can be used as a signature pattern for these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this pattern will pick up two of the four repeats in AdRab-B, the first one is not detected as its sequence has diverged in the region of the putative active site residue. The last one is also not detected because it is slightly divergent at the end of the pattern. Expert(s) to contact by email Upton C. upton@sol.uvic.ca Buckley J.T. tbuckley@sol.uvic.ca Last update November 1995 / First entry.

Pfam	Prosite	Full Name	Description
			References [1] Upton C., Buckley J.T. Trends Biochem. Sci. 20:178-179(1995).
Lipoprotein_1	PDOC00013	Prokaryotic membrane lipoprotein lipid attachment site	In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). - Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be. Last update November 1995 / Pattern and text revised. References [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990). [2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988). [3] von Heijne G. Protein Eng. 2:531-534(1989). [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).
Lipoprotein_2	PDOC00013	Prokaryotic membrane lipoprotein lipid attachment site	In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). - Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS. - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C

Pfam	Prosite	Full Name	Description
			(genes vlpABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification. Description of pattern(s) and/or profile(s) Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be. Last update November 1995 / Pattern and text revised. References [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990). [2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988). [3] von Heijne G. Protein Eng. 2:531-534(1989). [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).
Lipoprotein_5	PDOC00013	Prokaryotic membrane lipoprotein lipid attachment site	In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene

Pfam	Prosite	Full Name	Description
			lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). - Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS. - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification. Description of pattern(s) and/or profile(s) Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be. Last update November 1995 / Pattern and text revised. References

Pfam	Prosite	Full Name	Description
			[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990). [2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988). [3] von Heijne G. Protein Eng. 2:531-534(1989). [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).
Luteo_Vpg		Luteovirus putative VPg genome linked protein	Accession number: PF01659 Definition: Luteovirus putative VPg genome linked protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_970 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 191.70 191.70 Noise cutoffs: -47.90 -47.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94120742 Reference Title: Soybean dwarf luteovirus contains the third variant genome Reference Title: type in the luteovirus group. Reference Author: Rathjen JP, Karageorgos LE, Habili N, Waterhouse PM, Symons Reference Author: RH; Reference Location: Virology 1994;198:671-679. Database Reference INTERPRO; IPR001964; Comment: This family consists of several putative genome linked proteins. Comment: The genomic RNA of luteoviruses are linked to virally encoded genome Comment: proteins (VPg). Open reading frame 4 is thought to encode the VPg Comment: in Soybean dwarf luteovirus [1]. Comment: Luteoviruses have isometric capsids that contain a positive stand Comment: ssRNA genome, they have no DNA stage during their replication. Number of members: 32
MATH		MATH domain	Accession number: PF00917 Definition: MATH domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1602 (release 3.0) Gathering cutoffs: 17 0 Trusted cutoffs: 17.90 0.20 Noise cutoffs: 11.80 11.80 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96334294 Reference Title: TRAF proteins and meprins share a conserved domain. Reference Author: Uren AG, Vaux DL; Reference Location: Trends Biochem Sci 1996;21:244-245. Reference Number: [2] Reference Medline: 99342031 Reference Title: Crystallographic analysis of CD40 recognition and signaling Reference Title: by human TRAF2. Reference Author: McWhirter SM, Pullen SS, Holton JM, Crute JJ, Kehry MR, Reference Author: Alber T;

Pfam	Prosite	Full Name	Description
			Reference Location: Proc Natl Acad Sci U S A 1999;96:8408-8413. Reference Number: [3] Reference Medline: 99069615 Reference Title: Comparison of the complete protein sets of worm and yeast: Reference Title: orthology and divergence. Reference Author: Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Reference Author: Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Reference Author: Cherry JM, Botstein D; Reference Location: Science 1998;282:2022-2028. Database Reference: SCOP; 1qsc; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002083; Database Reference PDB; 1qsc A; 357; 498; Database Reference PDB; 1qsc B; 357; 498; Database Reference PDB; 1qsc C; 357; 498; Database reference: PFAMB; PB018448; Database reference: PFAMB; PB040690; Database reference: PFAMB; PB041198; Comment: This motif has been called the Meprin And TRAF-Homology Comment: (MATH) domain. This domain is hugely expanded in the nematode Comment: C. elegans [3]. Number of members: 212
MCT		Monocarboxylate transporter	Accession number: PF01587 Definition: Monocarboxylate transporter Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_483 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 322.90 322.90 Noise cutoffs: -38.20 -38.20 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98087501 Reference Title: Cloning and sequencing of four new mammalian Reference Title: monocarboxylate transporter (MCT) homologues confirms the Reference Title: existence of a transporter family with an ancient past. Reference Author: Price NT, Jackson VN, Halestrap AP, Reference Location: Biochem J 1998;329:321-328. Database Reference INTERPRO; IPR002897; Comment: This domain consists of the transmembrane region of the monocarboxylate Comment: transporters. Monocarboxylate transporters (MTC) are transmembrane Comment: glycoproteins with 10-12 predicted transmembrane regions. Comment: They catalyse the proton linked transport of lactic acid, Comment: pyruvate and ketone bodies across the plasma membrane [1]. Number of members: 33
Methionine_synt		Methionine synthase, vitamin-B12 independent	Accession number: PF01717 Definition: Methionine synthase, vitamin-B12 independent Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1909 (release 4.1) Gathering cutoffs: -155.0 -155.0 Trusted cutoffs: -155.00 -155.00 Noise cutoffs: -170.00 -170.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1]

Pfam	Prosite	Full Name	Description
			Reference Medline: 98301657 Reference Title: The specific features of methionine biosynthesis and Reference Title: metabolism in plants. Reference Author: Ravanel S, Gakiere B, Job D, Douce R; Reference Location: Proc Natl Acad Sci U S A 1998;95:7805-7812. Database Reference INTERPRO; IPR002629; Database reference: PFAMB; PB041617; Comment: This is a family of vitamin-B12 independent methionine synthases Comment: or 5-methyltetrahydropteroyltriglutamate--homocysteine Comment: methyltransferases, EC:2.1.1.14 from bacteria and plants. Comment: Plants are the only higher eukaryotes that have the required enzymes Comment: for methionine synthesis [1]. Comment: This enzyme catalyses the last step in the production of methionine Comment: by transferring a methyl group from 5-methyltetrahydrofolate to Comment: homocysteine [1]. Comment: The aligned region makes up the carboxy region of the approximately Comment: 750 amino acid protein except in some hypothetical archaeal proteins Comment: present in the family, where this region corresponds to the Comment: entire length. Number of members: 28
Methyltransf_2		O-methyltransferase	Accession number: PF00891 Definition: O-methyltransferase Previous Pfam IDs: Methyltransf; Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_152 (release 3.0) Gathering cutoffs: -53 -53 Trusted cutoffs: -22.00 -22.00 Noise cutoffs: -84.60 -84.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93167811 Reference Title: Purification of a 40-kilodalton methyltransferase active in Reference Title: the aflatoxin biosynthetic pathway. Reference Author: Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah Reference Author: AH; Reference Location: Appl Environ Microbiol 1993;59:479-484. Database Reference INTERPRO; IPR001077; Comment: This family includes a range of O-methyltransferases. These Comment: enzymes utilise S-adenosyl methionine. Number of members: 67
Methyltransf_3		O-methyltransferase	Accession number: PF01596 Definition: O-methyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_749 (release 4.1) Gathering cutoffs: -86 -86 Trusted cutoffs: -81.80 -81.80 Noise cutoffs: -91.00 -91.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97090395 Reference Title: Two multifunctional peptide synthetases and an Reference Title: O-methyltransferase are involved in the biosynthesis of the

Pfam	Prosite	Full Name	Description
			Reference Title: DNA-binding antibiotic and antitumour agent saframycin Mx1 Reference Title: from Myxococcus xanthus. Reference Author: Pospiech A, Bietenhader J, Schupp T; Reference Location: Microbiology 1996;142:741-746. Database Reference: SCOP; 1vid; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002935; Database Reference PDB; 1vid ; 13; 186; Database reference: PFAMB; PB040269; Comment: Members of this family are O-methyltransferases. The family Comment: includes catechol o-methyltransferase Swiss:P21964, caffeoyl-CoA Comment: O-methyltransferase Swiss:Q43095 and a family of bacterial Comment: O-methyltransferases that may be involved in antibiotic Comment: production [1]. Number of members: 39
MMR_HSR1		GTPase of unknown function	Accession number: PF01926 Definition: GTPase of unknown function Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -21 -21 Trusted cutoffs: -20.70 -20.70 Noise cutoffs: -31.60 -31.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94235953 Reference Title: Structure and evolution of a member of a new subfamily of Reference Title: GTP-binding proteins mapping to the human MHC class I Reference Title: region. Reference Author: Vernet C, Ribouchon MT, Chimini GPontarotti P; Reference Location: Mamm Genome 1994;5:100-105. Database Reference INTERPRO; IPR002917; Database reference: PFAMB; PB000471; Database reference: PFAMB; PB002171; Database reference: PFAMB; PB015790; Number of members: 67
MoaC		MoaC family	Accession number: PF01967 Definition: MoaC family Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 73.00 73.00 Noise cutoffs: -93.90 -93.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99337076 Reference Title: Characterization of a molybdenum cofactor biosynthetic gene Reference Title: cluster in Rhodobacter capsulatus which is specific for the Reference Title: biogenesis of dimethylsulfoxide reductase. Reference Author: Solomon PS, Shaw AL, Lane I, Hanson GR, Palmer T, McEwan Reference Author: AG; Reference Location: Microbiology 1999;145:1421-1429. Database Reference INTERPRO; IPR002820; Comment: Members of this family are involved in molybdenum Comment: cofactor biosynthesis. However their molecular Comment: function is not known.

Pfam	Prosite	Full Name	Description
			Number of members: 24
Morbilli_P		Morbillivirus RNA polymerase alpha subunit	Accession number: PF01647 Definition: Morbillivirus RNA polymerase alpha subunit Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_903 (release 4.1) Gathering cutoffs: -74 -74 Trusted cutoffs: 22.90 22.90 Noise cutoffs: -171.70 -171.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 92341068 Reference Title: Sequence analysis of the genes encoding the nucleocapsid Reference Title: protein and phosphoprotein (P) of phocid distemper virus, Reference Title: and editing of the P gene transcript. Reference Author: Blixenkrone-Moller M, Sharma B, Varsanyi TM, Hu A, Norrby Reference Author: E, Kovamees J; Reference Location: J. Gen. Virol. 1992;73:885-893. Database Reference INTERPRO; IPR002581; Database reference: PFAMB; PB002389; Comment: This family consists of morbillivirus RNA polymerase alpha subunit Comment: and non structural protein V. The P gene of morbillivirus is Comment: cotranscriptionally edited leading to the N-terminal Comment: half of the P protein being appended to the C-terminal of the P protein, Comment: and a cysteine rich region in the V fusion protein which has been Comment: shown to bind zinc [see Virology 3rd edition, volume 1, chapter 40, Comment: pages 1182-1184]. Comment: Morbilliviruses are positive strand ssRNA viruses and a part of the Comment: paramyxoviridae family, members include measles virus and phocine Comment: distemper virus. Number of members: 52
Myc_N_term		Myc amino-terminal region	Accession number: PF01056 Definition: Myc amino-terminal region Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_387 (release 3.0) Gathering cutoffs: -109 -109 Trusted cutoffs: -81.20 -81.20 Noise cutoffs: -137.40 -137.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98280742 Reference Title: The molecular role of Myc in growth and transformation: Reference Title: recent discoveries lead to new insights. Reference Author: Facchini LM, Penn LZ; Reference Location: FASEB J 1998;12:633-651. Reference Number: [2] Reference Medline: 97318600 Reference Title: Myc target genes. Reference Author: Grandori C, Eisenman RN; Reference Location: Trends Biochem Sci 1997;22:177-181. Database Reference INTERPRO; IPR002418; Comment: The myc family belongs to the basic helix-loop-helix leucine zipper Comment: class of transcription factors, see HLH. Myc forms a Comment: heterodimer with Max, and this complex regulates cell growth through

Pfam	Prosite	Full Name	Description
			Comment: direct activation of genes involved in cell replication [2]. Number of members: 56
Myosin_tail		Myosin tail	Accession number: PF01576 Definition: Myosin tail Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_356 (release 4.1) Gathering cutoffs: 19 19 Trusted cutoffs: 23.30 23.30 Noise cutoffs: 15.10 15.10 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 87060988 Reference Title: Complete nucleotide and encoded amino acid sequence of a Reference Title: mammalian myosin heavy chain gene. Evidence against Reference Title: intron-dependent evolution of the rod. Reference Author: Strehler EE, Strehler-page M-A, Perriard JC, Periasamy M, Reference Author: Nadal-ginard B; Reference Location: J MOL BIOL 1986;190:291-317. Database Reference INTERPRO; IPR002928; Comment: The myosin molecule is a multi-subunit complex made up Comment: of two heavy chains and four light chains it is a fundamental contractile Comment: protein found in all eukaryote cell types [1]. Comment: This family consists of the coiled-coil myosin heavy chain tail region. Comment: The coiled-coil is composed of the tail from two molecules of myosin. Comment: These can then assemble into the macromolecular thick filament [1]. Comment: The coiled-coil region provides the structural backbone the thick Comment: filament [1]. Number of members: 182
Na_Ala_symp	PDOC00681	Sodium:alanine symporter family signature	It has been shown [1] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:alanine symporter family (SAF) and currently consists of the following proteins: - Thermophilic bacterium PS-3 alanine carrier protein (ACP). ACP can use both sodium and hydrogen as a symport ion. - Alteromonas haloplanktis D-alanine/glycine permease (gene dagA). - Bacillus subtilis alsT. - Hypothetical protein yaaJ from Escherichia coli and HI0183, the corresponding Haemophilus influenzae protein. - Haemophilus influenzae hypothetical protein HI0883. These integral membrane proteins are predicted to comprise a least eight membrane spanning domains. As a signature pattern we selected a highly conserved region which is located in the N-terminal section and which includes part of the first transmembrane region.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern G-G-x-[GA](2)-[LIVM]-F-W-M-W-[LIVM]-x-[STAV]-[LIVMFA](2)-G Sequences known to belong to this class detected by ôhe pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).
Na_Ca_Ex		Sodium/calcium exchanger protein	Accession number: PF01699 Definition: Sodium/calcium exchanger protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1680 (release 4.1) Gathering cutoffs: 3 3 Trusted cutoffs: 3.40 3.40 Noise cutoffs: 1.20 1.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96394663 Reference Title: Cloning of a third mammalian Na+-Ca2+ exchanger, NCX3. Reference Author: Nicoll DA, Quednau BD, Qui Z, Xia YR, Lusis AJ, Philipson Reference Author: KD; Reference Location: J Biol Chem 1996;271:24914-24921. Reference Number: [2] Reference Medline: 91047958 Reference Title: Molecular cloning and functional expression of the cardiac Reference Title: sarcolemmal Na(+)-Ca2+ exchanger. Reference Author: Nicoll DA, Longoni S, Philipson KD; Reference Location: Science 1990;250:562-565. Database Reference INTERPRO; IPR002613; Database reference: PFAMB; PB002768; Database reference: PFAMB; PB040773; Database reference: PFAMB; PB041540; Comment: This is a family of sodium/calcium exchanger integral membrane Comment: proteins. This family covers the integral membrane regions of Comment: the proteins. Sodium/calcium exchangers regulate intracellular Ca2+ Comment: concentrations in many cells; cardiac myocytes, epithelial cells, Comment: neurons retinal rod photoreceptors and smooth muscle cells [2]. Comment: Ca2+ is moved into or out of the cytosol depending on Na+ concentration Comment: [2]. In humans and rats there are 3 isoforms; NCX1 NCX2 and NCX3 [1] Comment: see Swiss:Q01728, Swiss:P48768 and Swiss:P70549 respectively. Number of members: 105
Na_Galacto_symp	PDOC00680	Sodium:galactoside symporter family signature	It has been shown [1] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:galactoside symporter family (SGF) and currently consists of the following proteins:

Pfam	Prosite	Full Name	Description
			- The melibiose carrier (gene melB) from a variety of enterobacteria. This protein is responsible for melibiose transport and is capable of using hydrogen, sodium, and lithium cations as coupling cations for cotransport. - The lactose permease from Lactobacillus (gene lacS or lacY). This protein is responsible for the transport of beta-galactosides into the cell, with the concomitant export of a proton. It consists of two domains; a N-terminal SGF domain and a C-terminal domain that resembles that of enzyme IIA of the PEP:sugar phosphotransferase system. - The raffinose permease from Pediococcus pentosaceus. It also consists of a N-terminal SGF domain and a C-terminal IIA domain. - The glucuronide carrier (gene gusB or uidP) from Escherichia coli. - The xylose transporter (gene xylP) from Lactobacillus pentosus. - Escherichia coli hypothetical protein yagG. - Escherichia coli hypothetical protein yicJ. - Escherichia coli hypothetical protein yihO. - Escherichia coli hypothetical protein yihP. - Bacillus subtilis hypothetical protein yjmB. - Bacillus subtilis hypothetical protein ynaJ. Like sugar transport proteins, these integral membrane proteins are predicted to comprise twelve membrane spanning domains. As a signature pattern we selected a highly conserved region which is located in a cytoplasmic loop between the second and third transmembrane regions. This region starts with a conserved aspartate which has been shown [2], in melB, to be important for the activity of the protein. Description of pattern(s) and/or profile(s) Consensus pattern [DG]-x(3)-G-x(3)-[DN]-x(6,8)-[GA]-[KRHQ]-[FSA]-[KR]-[PT]- [FYW]-[LIVMWQ]-[LIV]-x-[GAFV]-[GSTA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994). [2] Pourcher T., Deckert M., Bassilana M., Leblanc G. Biochem. Biophys. Res. Commun. 178:1176-1181(1991).
Na_K_ATPase_C		Na+/K+ ATPase C-terminus	This domain is specific to the sodium and potassium ATPases (Na_K-ATPase). The sodium pump (Na+,K+ ATPase), located in the plasma membrane of all animal cells [1], is an heterotrimer of a catalytic subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta chain) and a small hydrophobic protein of about 6 Kd. The beta subunit seems [2] to regulate, through the assembly of alpha/beta heterodimers, the number of sodium pumps transported to the plasma membrane. This family is typically found in association with E1-E2 ATPase. Uses of these polypeptide includes regulating that ion content in a desired cell or organism and can convey salt or ion tolerance.
Na_K_ATPase_N		Na+/K+ ATPase C-	Accession number: PF00690

Pfam	Prosite	Full Name	Description
		terminus	Definition: Na+/K+ ATPase C-terminus Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_138 (release 2.1) Gathering cutoffs: 15.6 15.6 Trusted cutoffs: 15.60 15.60 Noise cutoffs: 15.10 15.10 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR000661; Database reference: PFAMB; PB000031; Comment: This family is always found in association with E1-E2_ATPase. Comment: This extension is specific to the Na+/K+ ATPase subfamily of Comment: ATPases. Number of members: 90
NAD_Gly3P_dh	PDOC00740	NAD-dependent glycerol-3-phosphate dehydrogenase signature	NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3- phosphate. It is a eukaryotic cytosolic homodimeric protein of about 40 Kd. As a signature pattern we selected a glycine-rich region that is probably [1] involved in NAD-binding. Description of pattern(s) and/or profile(s) Consensus pattern G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x- [DE]-G-[LIVM]-x-[LIVMFYW]-G-x-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Otto J., Argos P., Rossmann M.G. Eur. J. Biochem. 109:325-330(1980).
NifU_N		NifU-like N terminal domain	Accession number: PF01592 Definition: NifU-like N terminal domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_772 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 1.20 1.20 Noise cutoffs: -28.80 -28.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97032601 Reference Title: A modular domain of NifU, a nitrogen fixation cluster Reference Title: protein, is highly conserved in evolution. Reference Author: Hwang DM, Dempsey A, Tan KT, Liew CC; Reference Location: J Mol Evol 1996;43:536-540. Database Reference INTERPRO; IPR002871; Comment: This domain is found in NifU in combination with NifU-like. Comment: This domain is found on isolated in several bacterial species Comment: such as Swiss:O53156. The nif genes are responsible for nitrogen Comment: fixation. However this domain is found in bacteria that do not Comment: fix nitrogen, so it may have a broader significance in the cell Comment: than nitrogen fixation.

Pfam	Prosite	Full Name	Description
			Number of members: 32
NLPC_P60		NLP/P60 family	Accession number: PF00877 Definition: NLP/P60 family Author: Bateman A Alignment method of seed: HMM_built_from_alignment Source of seed members: Pfam-B_292 (release 3.0) Gathering cutoffs: -9 -9 Trusted cutoffs: -8.30 -8.30 Noise cutoffs: -10.40 -10.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR000064; Database reference: PFAMB; PB024706; Comment: The function of this domain is unknown. It is found Comment: in several lipoproteins. Number of members: 54
NTR		NTR/C345C module	Accession number: PF01759 Definition: NTR/C345C module Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 57.30 57.30 Noise cutoffs: 2.80 2.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99379676 Reference Title: The NTR module: domains of netrins, secreted frizzled Reference Title: related proteins, and type I procollagen C-proteinase Reference Title: enhancer protein are homologous with tissue inhibitors of Reference Title: metalloproteases [In Process Citation] Reference Author: Banyai L, Patthy L; Reference Location: Protein Sci 1999;8:1636-1642. Database Reference INTERPRO; IPR001134; Database reference: PFAMB; PB005955; Comment: We have not included the related TIMP family. Comment: It has been suggested that the common function of these Comment: modules is binding to metzincins [1]. A subset of this family Comment: is known as the C345C domain because it occurs in complement Comment: C3, C4 and C5. Number of members: 64
Nucleoside_tran		Nucleoside transporter	Accession number: PF01733 Definition: Nucleoside transporter Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_2135 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 25.50 25.50 Noise cutoffs: -122.50 -122.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98148080 Reference Title: Cloning of the human equilibrative, Reference Title: nitrobenzylmercaptopurine riboside (NBMPR)-insensitive Reference Title: nucleoside transporter ei by functional expression in a Reference Title: transport-deficient cell line. Reference Author: Crawford CR, Patel DH, Naeve C, Belt JA; Reference Location: J Biol Chem 1998;273:5288-5293. Reference Number: [2]

Pfam	Prosite	Full Name	Description
			Reference Medline: 98019212 Reference Title: Molecular cloning and functional characterization of Reference Title: nitrobenzylthioinosine (NBMPR)-sensitive (es) and Reference Title: NBMPR-insensitive (ei) equilibrative nucleoside transporter Reference Title: proteins (rENT1 and rENT2) from rat tissues. Reference Author: Yao SY, Ng AM, Muzyka WR, Griffiths M, Cass CE, Baldwin SA, Reference Author: Young JD; Reference Location: J Biol Chem 1997;272:28423-28430. Database Reference INTERPRO; IPR002259; Comment: This is a family of nucleoside transporters. Comment: In mammalian cells nucleoside transporters transport nucleoside Comment: across the plasma membrane and are essential for nucleotide Comment: synthesis via the salvage pathways for cells that lack their own Comment: de novo synthesis pathways [2]. Comment: Also in this family is mouse and human nucleolar protein HNP36 Comment: Swiss:Q14542 a protein of unknown function; although it has been Comment: hypothesized to be a plasma membrane nucleoside transporter [2]. Number of members: 15
Orbi_VP6		Orbivirus helicase VP6	Accession number: PF01516 Definition: Orbivirus helicase VP6 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_765 (release 4.0) Gathering cutoffs: -68 -68 Trusted cutoffs: -37.10 -37.10 Noise cutoffs: -98.90 -98.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97456481 Reference Title: Bluetongue virus VP6 protein binds ATP and exhibits an Reference Title: RNA-dependent ATPase function and a helicase activity that Reference Title: catalyze the unwinding of double-stranded RNA substrates. Reference Author: Stauber N, Martinez-Costas J, Sutton G, Monastyrskaya K, Reference Author: Roy P; Reference Location: J Virol 1997;71:7220-7226. Database Reference INTERPRO; IPR001399; Comment: The VP6 protein a minor protein in the core of the virion Comment: is probably the viral helicase [1]. Number of members: 27
OSCP	PDOC00327	ATP synthase delta (OSCP) subunit signature	ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1). One of the subunits of the ATPase complex, known as subunit delta in bacteria and chloroplasts or the Oligomycin Sensitivity Conferral Protein (OSCP) in mitochondria, seems to be part of the stalk that links CF(0) to CF(1). It

Pfam	Prosite	Full Name	Description
			either transmits conformational changes from CF(0) into CF(1) or is involved in proton conduction [3]. The different delta/OSCP subunits are proteins of approximately 200 amino-acid residues - once the transit peptide has been removed in the chloroplast and mitochondrial forms - which show only moderate sequence homology. The signature pattern used to detect ATPase delta/OSCP subunits is based on a conserved region in the C-terminal section of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x- [GSA]-G-[LIVMFYGA]-x-[LIVM]-[KRHENQ]-x-[GSEN] Sequences known to belong to this class detected by the pattern ALL, except 3 sequences. Other sequence(s) detected in SWISS-PROT 2. Last update November 1997 / Pattern and text revised. References [1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989). [2] Senior A.E. Physiol. Rev. 68:177-231(1988). [3] Engelbrecht S., Junge W. Biochim. Biophys. Acta 1015:379-390(1990).
OTCace	PDOC00091	Aspartate and ornithine carbamoyltransferases signature	Aspartate carbamoyltransferase (EC 2 1.3.2) (ATCase) catalyzes the conversion of aspartate and carbamoyl phosphate to carbamoylaspartate, the second step in the de novo biosynthesis of pyrimidine nucleotides [1]. In prokaryotes ATCase consists of two subunits: a catalytic chain (gene pyrB) and a regulatory chain (gene pyrl), while in eukaryotes it is a domain in a multi-functional enzyme (called URA2 in yeast, rudimentary in Drosophila, and CAD in mammals [2]) that also catalyzes other steps of the biosynthesis of pyrimidines. Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase) catalyzes the conversion of ornithine and carbamoyl phosphate to citrulline. In mammals this enzyme participates in the urea cycle [3] and is located in the mitochondrial matrix. In prokaryotes and eukaryotic microorganisms it is involved in the biosynthesis of arginine. In some bacterial species it is also involved in the degradation of arginine [4] (the arginine deaminase pathway). It has been shown [5] that these two enzymes are evolutionary related. The predicted secondary structure of both enzymes are similar and there are some regions of sequence similarities. One of these regions includes three

Pfam	Prosite	Full Name	Description
			residues which have been shown, by crystallographic studies [6], to be implicated in binding the phosphoryl group of carbamoyl phosphate. We have selected this region as a signature for these enzymes. Description of pattern(s) and/or profile(s) Consensus pattern F-x-[EK]-x-S-[GT]-R-T [S, R, and the 2nd T bind carbamoyl phosphate] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note the residue in position 3 of the pattern allows to distinguish between an ATCase (Glu) and an OTCase (Lys). Last update October 1993 / Text revised. References [1] Lerner C.G., Switzer R.L. J. Biol. Chem. 261:11156-11165(1986). [2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993). [3] Takiguchi M., Matsubasa T., Amaya Y., Mori M. BioEssays 10:163-166(1989). [4] Baur H., Stalon V., Falmagne P., Luethi E., Haas D. Eur. J. Biochem. 166:111-117(1987). [5] Houghton J.E., Bencini D.A., O'Donovan G.A., Wild J.R. Proc. Natl. Acad. Sci. U.S.A. 81:4864-4868(1981). [6] Ke H.-M., Honzatko R.B., Lipscomb W.N. Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).
oxidored_q1_N		NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus	Accession number: PF00662 Definition: NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_22 (release 2.1) Gathering cutoffs: 18 18 Trusted cutoffs: 19.40 19.40 Noise cutoffs: 16.70 16.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93110040 Reference Title: The NADH:ubiquinone oxidoreductase (complex I) of Reference Title: respiratory chains. Reference Author: Walker JE; Reference Location: Q Rev Biophys 1992;25:253-324. Database Reference INTERPRO; IPR001516; Database reference: PFAMB; PB000410; Database reference: PFAMB; PB033295; Database reference: PFAMB; PB040550; Comment: This sub-family represents an amino terminal extension Comment: of oxidored_q1. Only NADH-Ubiquinone chain 5 and Comment: eubacterial chain L are in this family. Comment: This sub-family is part of complex I which

Pfam	Prosite	Full Name	Description
			catalyses the Comment: transfer of two electrons from NADH to ubiquinone in a Comment: reaction that is associated with proton translocation Comment: across the membrane. Number of members: 546
oxidored_q2		NADH-ubiquinone/plastoquinone oxidoreductase chain 4L	Accession number: PF00420 Definition: NADH-ubiquinone/plastoquinone oxidoreductase chain 4L Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_193 (release 1.0) Gathering cutoffs: 25 15 Trusted cutoffs: 29.70 29.70 Noise cutoffs: 20.40 20.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR001133; Database reference: PFAMB; PB006066; Number of members: 219
PAN	PDOC00376	Apple domain	Plasma kallikrein (EC 3.4.21.34) and coagulation factor XI (EC 3.4.21.27) are two related plasma serine proteases activated by factor XIIA and which share the same domain topology: an N-terminal region that contains four tandem repeats of about 90 amino acids and a C-terminal catalytic domain. The 90 amino-acid repeated domain contains 6 conserved cysteines. It has been shown [1,2] that three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines. The domain can be drawn in the shape of an apple (see below) and has been accordingly called the 'apple domain'. xxx xxx x C--C x x x x x x Cxxx x x | x x x Schematic representation of an x Cxxx x apple domain. x x x x x x x x x xxx x x x xx xx C---C .....x x..... Apart from the cysteines, there are a number of other conserved positions in the apple domain. We have developed a pattern, that spans the complete domain, and which includes these conserved positions. Description of pattern(s) and/or profile(s) Consensus pattern C-x(3)-[LIVMFY]-x(5)-[LIVMFY]-x(3)-[DENQ]-[LIVMFY]-x(10)- C-x(3)-C-T-x(4)-C-x-[LIVMFY]-F-x-[FY]-x(13,14)-C-x- [LIVMFY]-[RK]-x-[ST]-x(14,15)-S-G-x-[ST]-[LIVMFY]-x(2)-C Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update June 1992 / Pattern and text revised. References

Pfam	Prosite	Full Name	Description
			[1] McMullen B.A., Fujikawa K., Davie E.W. Biochemistry 30:2050-2056(1991). [2] McMullen B.A., Fujikawa K., Davie E.W. Biochemistry 30:2056-2060(1991).
PAP2		PAP2 superfamily	Accession number: PF01569 Definition: PAP2 superfamily Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_486 (release 4.0) Gathering cutoffs: 16 16 Trusted cutoffs: 22.00 22.00 Noise cutoffs: 11.40 11.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97194074 Reference Title: Identification of a novel phosphatase sequence motif. Reference Author: Stukey J, Carman GM; Reference Location: Protein Sci 1997;6:469-472. Reference Number: [2] Reference Medline: 97406916 Reference Title: An unexpected structural relationship between integral Reference Title: membrane phosphatases and soluble haloperoxidases. Reference Author: Neuwald AF; Reference Location: Protein Sci 1997;6:1764-1767. Database Reference INTERPRO; IPR000326; Database reference: PFAMB; PB021113; Database reference: PFAMB; PB040926; Database reference: PFAMB; PB041096; Database reference: PFAMB; PB041301; Comment: This family includes the enzyme type 2 phosphatidic acid Comment: phosphatase (PAP2). Number of members: 49
PAPS_reduct		Phosphoadenosine phosphosulfate reductase family	Accession number: PF01507 Definition: Phosphoadenosine phosphosulfate reductase family Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_590 (release 4.0) Gathering cutoffs: 49 49 Trusted cutoffs: 55.40 55.40 Noise cutoffs: -34.60 -34.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97411695 Reference Title: Crystal structure of phosphoadenylyl sulphate (PAPS) Reference Title: reductase: a new family of adenine nucleotide alpha Reference Title: hydrolases. Reference Author: Savage H, Montoya G, Svensson C, Schwenn JD, Sinning I; Reference Location: Structure 1997;5:895-906. Reference Number: [2] Reference Medline: 96061968 Reference Title: Reaction mechanism of thioredoxin: Reference Title: 3'-phospho-adenylylsulfate reductase investigated by Reference Title: site-directed mutagenesis. Reference Author: Berendt U, Haverkamp T, Prior A, Schwenn JD; Reference Location: Eur J Biochem 1995;233:347-356. Reference Number: [3] Reference Medline: 91066949

Pfam	Prosite	Full Name	Description
			Reference Title: ATP sulphurylase activity of the nodP and nodQ gene Reference Title: products of Rhizobium meliloti. Reference Author: Schwedock J, Long SR; Reference Location: Nature 1990;348:644-647. Database Reference: SCOP; 1sur; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002500; Database Reference PDB; 1sur ; 48; 215; Comment: This domain is found in phosphoadenosine phosphosulfate (PAPS) reductase Comment: enzymes or PAPS sulfotransferase. PAPS reductase is part of the adenine Comment: nucleotide alpha hydrolases superfamily also including N type ATP PPases Comment: and ATP sulphurylases [1]. The enzyme uses thioredoxin as an electron Comment: donor for the reduction of PAPS to phospho-adenosine-phosphate (PAP) [1,2]. Comment: It is also found in NodP nodulation protein P from Rizobium which has ATP Comment: sulpurylase activity (sulfate adenylate transferase) [3]. Number of members: 48
PARP		Poly(ADP-ribose) polymerase catalytic region	Accession number: PF00644 Definition: Poly(ADP-ribose) polymerase catalytic region. Author: Bateman A Alignment method of seed: HMM_built_from_alignment Source of seed members: Bateman A Gathering cutoffs: -59.4 -59.4 Trusted cutoffs: -44.60 -44.60 Noise cutoffs: -180.60 -180.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96353841 Reference Title: Structure of the catalytic fragment of poly(AD-ribose) Reference Title: polymerase from chicken. Reference Author: Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE; Reference Location: Proc Natl Acad Sci U S A 1996;93:7481-7485. Reference Number: [2] Reference Medline: 93293867 Reference Title: The carboxyl-terminal domain of human poly(ADP-ribose) Reference Title: polymerase. Overproduction in Escherichia coli, large scale Reference Title: purification, and characterization. Reference Author: Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G, Reference Author: Althaus FR; Reference Location: J Biol Chem 1993;268:13454-13461. Database Reference: SCOP; 1paw; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001290; Database Reference PDB; 1a26 ; 662; 997; Database Reference PDB; 1pax ; 662; 997; Database Reference PDB; 2pax ; 662; 997; Database Reference PDB; 3pax ; 662; 997; Database Reference PDB; 4pax ; 662; 997; Database Reference PDB; 2paw ; 662; 1009; Database reference: PFAMB; PB041409; Comment: Poly(ADP-ribose) polymerase catalyses the covalent Comment: attachment of ADP-ribose units from NAD+ to itself and Comment: to a limited number of other DNA binding proteins, which Comment: decreases their affinity for DNA. Comment: Poly(ADP-ribose) polymerase is a regulatory component

Pfam	Prosite	Full Name	Description
			Comment: induced by DNA damage. Comment: The carboxyl-terminal region is the most highly conserved Comment: region of the protein. Experiments have shown that a Comment: carboxyl 40 kDa fragment is still catalytically active [2]. Number of members: 19
PC_rep		Proteasome/cyclosome repeat	Accession number: PF01851 Definition: Proteasome/cyclosome repeat Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 0 Trusted cutoffs: 30.60 3.00 Noise cutoffs: 15.80 15.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97348748 Reference Title: A repetitive sequence in subunits of the 26S proteasome and Reference Title: 20S cyclosome (anaphase-promoting complex). Reference Author: Lupas A, Baumeister W, Hofmann K; Reference Location: Trends Biochem Sci 1997;22:195-196. Database Reference INTERPRO; IPR002015; Database reference: PFAMB; PB009978; Database reference: PFAMB; PB040656; Number of members: 112
PE		PE family	Accession number: PF00934 Definition: PE family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_253 (release 3.0) Gathering cutoffs: -20 -20 Trusted cutoffs: -10.80 -10.80 Noise cutoffs: -20.60 -20.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98295987 Reference Title: Deciphering the biology of Mycobacterium tuberculosis from Reference Title: the complete genome sequence. Reference Author: Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Reference Author: Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Reference Author: Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Reference Author: Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin Reference Author: N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al; Reference Location: Nature 1998;393:537-544. Database Reference INTERPRO; IPR000084; Comment: This family named after a PE motif near to the amino Comment: terminus of the domain. The PE family of proteins Comment: all contain an amino-terminal region of about 110 Comment: amino acids. The carboxyl terminus of this family Comment: are variable and fall into several classes. The Comment: largest class of PE proteins is the highly repetitive Comment: PGRS class which have a high glycine content. Comment: The function of these proteins is uncertain

Pfam	Prosite	Full Name	Description
			but it Comment: has been suggested that they may be related to Comment: antigenic variation of Mycobacterium tuberculosis [1]. Number of members: 90
Pep_deformylase		Polypeptide deformylase	Accession number: PF01327 Definition: Polypeptide deformylase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Sarah Teichmann Gathering cutoffs: 25 25 Trusted cutoffs: 157.40 157.40 Noise cutoffs: -29.00 -29.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97002011 Reference Title: A new subclass of the zinc metalloproteases superfamily Reference Title: revealed by the solution structure of peptide deformylase. Reference Author: Meinnel T, Blanquet S, Dardel F; Reference Location: J Mol Biol 1996;262:375-386. Reference Number: [2] Reference Medline: 98332750 Reference Title: Solution structure of nickel-peptide deformylase. Reference Author: Dardel F, Ragusa S, Lazennec C, Blanquet S, Meinnel T; Reference Location: J Mol Biol 1998;280:501-513. Database Reference: SCOP; 1def; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000181; Database Reference PDB; 2def ; 4; 142; Database Reference PDB; 1def ; 4; 142; Database Reference PDB; 1dff ; 4; 142; Database Reference PDB; 1bsj A; 4; 142; Database Reference PDB; 1bsk A; 4; 142; Database Reference PDB; 1bs4 A; 4; 142; Database Reference PDB; 1bs4 B; 504; 642; Database Reference PDB; 1bs4 C; 1004; 1142; Database Reference PDB; 1bs5 A; 4; 142; Database Reference PDB; 1bs5 B; 504; 642; Database Reference PDB; 1bs5 C; 1004; 1142; Database Reference PDB; 1bs6 A; 4; 142, Database Reference PDB; 1bs6 B; 504; 642; Database Reference PDB; 1bs6 C; 1004; 1142; Database Reference PDB; 1bs7 A; 4; 142; Database Reference PDB; 1bs7 B; 504; 642; Database Reference PDB; 1bs7 C; 1004; 1142; Database Reference PDB; 1bs8 A; 4; 142; Database Reference PDB; 1bs8 B; 504; 642; Database Reference PDB; 1bs8 C; 1004; 1142; Database Reference PDB; 1bsz A; 4; 142; Database Reference PDB; 1bsz B; 504; 642; Database Reference PDB; 1bsz C; 1004; 1142; Database Reference PDB; 1icj A; 4; 142; Database Reference PDB; 1icj B; 504; 642; Database Reference PDB; 1icj C; 1004; 1142; Database reference: PFAMB; PB041251; Number of members: 25
Peptidase_C15		Pyroglutamyl peptidase	Accession number: PF01470 Definition: Pyroglutamyl peptidase Author: Bateman A Alignment method of seed: Clustalw_manual Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 436.10 436.10 Noise cutoffs: -155.40 -155.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM

Pfam	Prosite	Full Name	Description
			Reference Number: [1] Reference Medline: 99216536 Reference Title: The crystal structure of pyroglutamyl peptidase I from Reference Title: bacillus amyloliquefaciens reveals a new structure for a Reference Title: cysteine protease. Reference Author: Odagaki Y, Hayashi A, Okada K, Hirotsu K, Kabashima T, Ito Reference Author: K, Yoshimoto T, Tsuru D, Sato M, Clardy J Reference Location: Structure 1999;7:399-411. Database Reference: SCOP; 1aug; fa; [SCOP-USA][CATH-PDBSUM] Database Reference MEROPS; C15; Database Reference INTERPRO; IPR000816; Database Reference PDB; 1a2z A; 2; 209; Database Reference PDB; 1a2z B; 2; 209; Database Reference PDB; 1a2z C; 2; 209; Database Reference PDB; 1a2z D; 2; 209; Database Reference PDB; 1aug A; 3; 204; Database Reference PDB; 1aug B; 213; 414; Database Reference PDB; 1aug C; 423; 624; Database Reference PDB; 1aug D; 633; 834; Number of members: 10
Peptidase_M20	PDOC00613	ArgE / dapE / ACY1 / CPG2 / yscS family signatures	The following enzymes have been shown [1,2,3] to be evolutionary and Functionally related: - In the biosynthetic pathway from glutamate to arginine, the removal of an acetyl group from N2-acetylornithine can be catalyzed via two distinct enzymatic strategies depending on the organism. In some bacteria and in fungi, the acetyl group is transferred on glutamate by glutamate acetyltransferase (EC 2.3.1.35) while in enterobacteria such as Escherichia coli, it is hydrolyzed by acetylornithine deacetylase (EC 3.5.1.16) (acetylornithinase) (AO) (gene argE). AO is a homodimeric cobalt-dependent enzyme which displays broad specificity and can also deacylates substrates such as acetylarginine, acetylhistidine, acetylglutamate semialdehyde, etc. - Succinyldiaminopimelate desuccinylase (EC 3.5.1.18) (SDAP) (gene dapE) is the enzyme which catalyzes the fifth step in the biosynthesis of lysine from aspartate semialdehyde: the hydrolysis of succinyl-diaminopimelate to diaminopimelate and succinate. SDAP is an enzyme that requires cobalt or zinc as a cofactor. - Aminoacylase-1 [4] (EC 3.5.1.14) (N-acyl-l-amino-acid amidohydrolase) (ACY1). ACY1 is a homodimeric zinc-binding mammalian enzyme that catalyzes the hydrolysis of N-alpha-acylated amino acids (except for aspartate). - Carboxypeptidase G2 (EC 3.4.17.11) (folate hydrolase G2) (gene cpg2) from Pseudomonas strain RS-16. This enzyme catalyzes the hydrolysis of reduced and non-reduced folates to pteroates and glutamate. G2 is a homodimeric zinc-dependent enzyme. - Vacuolar carboxypeptidase S (EC 3.4.17.4) (yscS) from yeast (gene CPS1). - Peptidase T (EC 3.4.11.-) (gene pepT) (tripeptidase) from bacteria. This

Pfam	Prosite	Full Name	Description
			enzyme catalyzes a variety of tripeptides containing N-terminal methionine, leucine, or phenylalanine. - Xaa-His dipeptidase (EC 3.4.13.3) (carnosinase) from Lactobacillus (gene pepV) [5], a metalloenzyme with activity against beta-alanyl-dipeptides including carnosine (beta-alanyl-histidine). These enzymes share a few characteristics. They hydrolyse peptidic bonds in Substrates that share a common structure, they are dependent on cobalt or zinc For their activity and they are proteins of 40 Kd to 60 Kd with a number of Regions of sequence similarity. As signature patterns for these proteins, we selected two of the conserved Regions. The first pattern contains a conserved histidine which could be Involved in binding metal ions and the second pattern contains a number of Conserved charged residues. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-[GALMY]-[LIVMF]-x-[GSA]-H-x-D-[TV]-[STAV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 6. Consensus pattern [GSTAI]-[SANQ]-D-x-K-[GSACN]-x(2)-[LIVMA]-x(2)-[LIVMFY]- x(14,17)-[LIVM]-x-[LIVMF]-[LIVMSTAG]-[LIVMFA]-x(2)-[DNG]- E-E-x-[GSTN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Note these proteins belong to families M20A/M20B in the classification of peptidases [6,E1]. Last update November 1997 / Patterns and text revised. References [1] Meinnel T., Schmitt E., Mechulam Y., Blanquet S. J. Bacteriol. 174:2323-2331(1992). [2] Boyen A., Charlier D., Sakanyan V., Mett I., Glansdorff N. Gene 116:1-6(1992). [3] Miller C G., Miller J.L., Bagga D A. J. Bacteriol. 173:3554-3558(1991). [4] Mitta M., Ohnogi H., Yamamoto A., Kato I., Sakiyama F., Tsunasawa S. J. Biochem. 112:737-742(1992). [5] Vongerichten K., Klein J., Matern H., Plapp R. Microbiology 140:2591-2600(1994). [6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt

Pfam	Prosite	Full Name	Description
Peptidase_M3	PDOC00129	Neutral zinc metallopeptidases, zinc-binding region signature	The majority of zinc-dependent metallopeptidases (with the notable exception Of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily,known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families. Family M1 - Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN). - Mammalian aminopeptidase N (EC 3.4.11.2). - Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells. - Yeast aminopeptidase yscII (gene APE2). - Yeast alanine/arginine aminopeptidase (gene AAP1). - Yeast hypothetical protein YIL137c. - Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity. Family M2 - Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers. Family M3 - Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides. - Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase). - Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion. - Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD). - Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp). - Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC). - Yeast hypothetical protein YKL134c. Family M4 - Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus. - Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB). - Extracellular elastase from Staphylococcus epidermidis. - Extracellular protease prt1 from Erwinia carotovora. - Extracellular minor protease smp from Serratia marcescens. - Vibriolysin (EC 3.4.24.25) from various species of Vibrio. - Protease prtA from Listeria monocytogenes.

Pfam	Prosite	Full Name	Description
			- Extracellular proteinase proA from Legionella pneumophila. Family M5 - Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi. Family M6 - Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins. Family M7 - Streptomyces extracellular small neutral proteases Family M8 - Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of Leishmania. Family M9 - Microbial collagenase (EC 3.4.24.3) from Clostridium perfringens and Vibrio alginolyticus. Family M10A - Serralysin (EC 3.4.24.40), an extracellular metalloprotease from Serratia. - Alkaline metalloproteinase from Pseudomonas aeruginosa (gene aprA). - Secreted proteases A, B, C and G from Erwinia chrysanthemi. - Yeast hypothetical protein YIL108w. Family M10B - Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase). - Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that allows the embryo to digest the protective envelope derived from the egg extracellular matrix. - Soybean metalloendoproteinase 1. Family M11 - Chlamydomonas reinhardtii gamete lytic enzyme (GLE). Family M12A - Astacin (EC 3.4.24.21), a crayfish endoprotease. - Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border metalloendopeptidase. - Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The Drosophila homolog of BMP-1 is the dorsal-ventral patterning protein tolloid. - Blastula protease 10 (BP10) from Paracentrotus lividus and the related protein SpAN from Strongylocentrotus purpuratus. - Caenorhabditis elegans protein toh-2. - Caenorhabditis elegans hypothetical protein F42A10.8. - Choriolysins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish Oryzias lapides.

Pfam	Prosite	Full Name	Description
			These proteases participates in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching. Family M12B - Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53). - Mouse cell surface antigen MS2. Family M13 - Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP). - Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide. - Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase. - Peptidase O from Lactococcus lactis (gene pepO). Family M27 - Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8]. Family M30 - Staphylococcus hyicus neutral metalloprotease. Family M32 - Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from Thermus aquaticus which is most active at high temperature. Family M34 - Lethal factor (LF) from Bacillus anthracis, one of the three proteins composing the anthrax toxin. Family M35 - Deuterolysin (EC 3.4.24.39) from Penicillium citrinum and related proteases from various species of Aspergillus. Family M36 - Extracellular elastinolytic metalloproteinases from Aspergillus. From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern [GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ] [The two H's are zinc ligands] [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for members of families M5, M7 amd M11. Other sequence(s) detected in SWISS-PROT 57; including Neurospora crassa conidiation-specific protein 13 which could be a zinc-protease. Last update July 1999 / Text revised. References [1] Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989) [2] Murphy G.J.P., Murphy G., Reynolds J.J. FEBS Lett. 289:4-7(1991). [3] Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B., Stoecker W. Zoology 99:237-246(1996). [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995). [5] Woessner J. Jr. FASEB J. 5:2145-2154(1991). [6] Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992). [7] Montecucco C., Schiavo G. Trends Biochem. Sci. 18:324-327(1993). [8] Niemann H., Blasi J., Jahn R. Trends Cell Biol. 4:179-185(1994). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_M48		Peptidase family M48	Accession number: PF01435 Definition: Peptidase family M48 Author: Bateman A Alignment method of seed: Clustalw_manual Source of seed members: Swiss-Prot Gathering cutoffs: -35 -35 Trusted cutoffs: -34.00 -34.00 Noise cutoffs: -42.20 -42.20 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference MEROPS; M48; Database Reference INTERPRO; IPR001915; Database reference: PFAMB; PB008839; Database reference: PFAMB; PB041497; Number of members: 28
Peptidase_S24		Peptidase family S24	Accession number: PF00717 Definition: Peptidase family S24 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_616 (release 2.1) Gathering cutoffs: 1 1 Trusted cutoffs: 2.00 2.00 Noise cutoffs: -9.00 -9.00

Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference MEROPS; S24; Database Reference: SCOP; 1umu; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000129; Database Reference PDB; 1adr ; 72; 76; Database Reference PDB; 1lmb 3; 89; 92; Database Reference PDB; 1lmb 4; 89; 92; Database Reference PDB; 1leb ; 66; 72; Database Reference PDB; 1umu A; 32; 123; Database Reference PDB; 1umu B; 32; 123; Database Reference PDB; 1ay9 A; 32; 123; Database Reference PDB; 1ay9 B; 32; 123; Database reference: PFAMB; PB005958; Database reference: PFAMB; PB041146; Database reference: PFAMB; PB041823; Number of members: 42
Peptidase_S8	PDOC00125	Serine proteases, subtilase family, active sites	Subtilases [1,2] are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases and can be used as signatures specific to that category of proteases. The subtilase family currently includes the following proteases: - Subtilisins (EC 3.4.21.62), these alkaline proteases from various Bacillus species have been the target of numerous studies in the past thirty years. - Alkaline elastase YaB from Bacillus sp. (gene ale). - Alkaline serine exoprotease A from Vibrio alginolyticus (gene proA). - Aqualysin I from Thermus aquaticus (gene pstI). - AspA from Aeromonas salmonicida. - Bacillopeptidase F (esterase) from Bacillus subtilis (gene bpf). - C5A peptidase from Streptococcus pyogenes (gene scpA). - Cell envelope-located proteases PI, PII, and PIII from Lactococcus lactis. - Extracellular serine protease from Serratia marcescens. - Extracellular protease from Xanthomonas campestris. - Intracellular serine protease (ISP) from various Bacillus. - Minor extracellular serine protease epr from Bacillus subtilis (gene epr). - Minor extracellular serine protease vpr from Bacillus subtilis (gene vpr). - Nisin leader peptide processing protease nisP from Lactococcus lactis. - Serotype-specific antigene 1 from Pasteurella haemolytica (gene ssa1). - Thermitase (EC 3.4.21.66) from Thermoactinomyces vulgaris. - Calcium-dependent protease from Anabaena variabilis (gene prcA). - Halolysin from halophilic bacteria sp. 172p1 (gene hly). - Alkaline extracellular protease (AEP) from Yarrowia lipolytica (gene xpr2). - Alkaline proteinase from Cephalosporium acremonium (gene alp). - Cerevisin (EC 3.4.21.48) (vacuolar protease B) from yeast (gene PRB1). - Cuticle-degrading protease (pr1) from Metarhizium anisopliae. - KEX-1 protease from Kluyveromyces lactis.

Pfam	Prosite	Full Name	Description
			- Kexin (EC 3.4.21.61) from yeast (gene KEX-2). - Oryzin (EC 3.4.21.63) (alkaline proteinase) from Aspergillus (gene alp). - Proteinase K (EC 3.4.21.64) from Tritirachium album (gene proK). - Proteinase R from Tritirachium album (gene proR). - Proteinase T from Tritirachium album (gene proT). - Subtilisin-like protease III from yeast (gene YSP3). - Thermomycolin (EC 3.4.21.65) from Malbranchea sulfurea. - Furin (EC 3.4.21.85), neuroendocrine convertases 1 to 3 (NEC-1 to -3) and PACE4 protease from mammals, other vertebrates, and invertebrates. These proteases are involved in the processing of hormone precursors at sites comprised of pairs of basic amino acid residues [3]. - Tripeptidyl-peptidase II (EC 3.4.14.10) (tripeptidyl aminopeptidase) from Human. - Prestalk-specific proteins tagB and tagC from slime mold [4]. Both proteins consist of two domains: a N-terminal subtilase catalytic domain and a C- terminal ABC transporter domain (see <PDOC00185>). Description of pattern(s) and/or profile(s) Consensus pattern [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH] [D is the active site residue] Sequences known to belong to this class detected by the pattern the majority of subtilases with a few exceptions. Other sequence(s) detected in SWISS-PROT 44. Consensus pattern H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM] [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for aspA and ssa1 which both seem to lack the histidine active site. Other sequence(s) detected in SWISS-PROT adenylate cyclase type VIII. Consensus pattern G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG] [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for nisP, tagC and S.marcescens extracellular serine protease. Other sequence(s) detected in SWISS-PROT 6. Note if a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100% Note these proteins belong to family S8 in the classification of peptidases [5,E1]. Expert(s) to contact by email Brannigan J. jab5@vaxa.york.ac.uk Siezen R.J. siezen@nizo.nl Last update November 1997 / Patterns and text revised. References [1] Siezen R.J., de Vos W.M., Leunissen J.A.M., Dijkstra B.W. Protein Eng. 4:719-737(1991). [2] Siezen R.J. (In) Proceeding subtilisin symposium, Hamburg, (1992). [3]

Pfam	Prosite	Full Name	Description
			Barr P.J. Cell 66:1-3(1991). [4] Shaulsky G., Kuspa A., Loomis W.F.; Genes Dev. 9:1111-1122(1995). [5] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_S9	PDOC00587	Prolyl oligopeptidase family serine active site	The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below. - Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences. - Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and argininyl residues. - Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. - Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor. - Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2). - Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus. A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A. Other sequence(s) detected in SWISS-PROT NONE. Note these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1]. Last update November 1997 / Text revised. References [1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991). [2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992). [3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992). [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
Peptidase_U7		Peptidase family U7	Accession number: PF01343 Definition: Peptidase family U7 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_707 (release 2.1) Gathering cutoffs: 25 25 Trusted cutoffs: 47.60 47.60 Noise cutoffs: -55.60 -55.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference MEROPS; U7; Database Reference INTERPRO; IPR002142; Number of members: 37
PEP-utilizers	PDOC00527	PEP-utilizing enzymes signatures	A number of enzymes that catalyze the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) via a phospho-histidine intermediate have been shown to be structurally related [1,2,3,4]. These enzymes are: - Pyruvate,orthophosphate dikinase (EC 2.7.9.1) (PPDK). PPDK catalyzes the reversible phosphorylation of pyruvate and phosphate by ATP to PEP and diphosphate. In plants PPDK function in the direction of the formation of PEP, which is the primary acceptor of carbon dioxide in C4 and crassulacean acid metabolism plants. In some bacteria, such as Bacteroides symbiosus, PPDK functions in the direction of ATP synthesis. - Phosphoenolpyruvate synthase (EC 2.7.9.2) (pyruvate,water dikinase). This enzyme catalyzes the reversible phosphorylation of pyruvate by ATP to form PEP, AMP and phosphate, an essential step in gluconeogenesis when pyruvate and lactate are used as a carbon source. - Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9). This is the first enzyme of the phosphoenolpyruvate-dependent sugar phosphotransferase

Pfam	Prosite	Full Name	Description
			system (PTS), a major carbohydrate transport system in bacteria. The PTS catalyzes the phosphorylation of incoming sugar substrates concomitant with their translocation across the cell membrane. The general mechanism of the PTS is the following: a phosphoryl group from PEP is transferred to enzyme-I (EI) of PTS which in turn transfers it to a phosphoryl carrier protein (HPr). Phospho-HPr then transfers the phosphoryl group to a sugar-specific permease. All these enzymes share the same catalytic mechanism: they bind PEP and transfer the phosphoryl group from it to a histidine residue. The sequence around that residue is highly conserved and can be used as a signature pattern for these enzymes. As a second signature pattern we selected a conserved region in the C-terminal part of the PEP-utilizing enzymes. The biological significance of this region is not yet known. Description of pattern(s) and/or profile(s) Consensus pattern G-[GA]-x-[STN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG] [H is phosphorylated] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [DEQSK]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q- [LIVMFYGT]-[STALIV]-[LIVMFY]-[GAS]-x(2)-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Patterns and text revised. References [1] Reizer J., Hoischen C., Reizer A., Pham T.N., Saier M.H. Jr. Protein Sci. 2:506-521(1993). [2] Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J., Saier M.H. Jr. Gene 181:103-108(1996). [3] Pocalyko D.J., Carroll L.J., Martin B.M., Babbitt P.C., Dunaway-Mariano D. Biochemistry 29:10757-10765(1990). [4] Niersbach M., Kreuzaler F , Geerse R.H., Postma P., Hirsch H.J. Mol. Gen. Genet. 232:332-336(1992).
PG_binding_2		Putative peptidoglycan binding domain	Accession number: PF01476 Definition: Putative peptidoglycan binding domain Author: Bateman A Alignment method of seed: HMM_built_from_alignment Source of seed members: Bateman A Gathering cutoffs: 22 22 Trusted cutoffs: 22.40 22.10 Noise cutoffs: 21.10 21.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 92324582

Pfam	Prosite	Full Name	Description
			Reference Title: Modular design of the Enterococcus hirae muramidase-2 and Reference Title: Streptococcus faecalis autolysin. Reference Author: Joris B, Englebert S, Chu CP, Kariyama R, Daneo-Moore L, Reference Author: Shockman GD, Ghuysen JM; Reference Location: FEMS Microbiol Lett 1992;70:257-264. Database Reference INTERPRO; IPR002482; Database reference: PFAMB; PB019287; Database reference: PFAMB; PB040847; Database reference: PFAMB; PB040977; Comment: This domain is about 40 residues long. It is found in a variety Comment: of enzymes involved in bacterial cell wall degradation [1]. This Comment: domain may have a general peptidoglycan binding function. Number of members: 197
phoslip	PDOC00109	Phospholipase A2 active sites signatures	Phospholipase A2 (EC 3.1.1.4) (PA2) [1,2] is an enzyme which releases fatty acids from the second carbon group of glycerol. PA2's are small and rigid proteins of 120 amino-acid residues that have four to seven disulfide bonds. PA2 binds a calcium ion which is required for activity. The side chains of two conserved residues, a histidine and an aspartic acid, participate in a 'catalytic network'. Many PA2's have been sequenced from snakes, lizards, bees and mammals. In the latter, there are at least four forms: pancreatic, membrane-associated as well as two less characterized forms. The venom of most snakes contains multiple forms of PA2. Some of them are presynaptic neurotoxins which inhibit neuromuscular transmission by blocking acetylcholine release from the nerve termini. We derived two different signature patterns for PA2's. The first is centered on the active site histidine and contains three cysteines involved in disulfide bonds. The second is centered on the active site aspartic acid and also contains three cysteines involved in disulfide bonds. Description of pattern(s) and/or profile(s) Consensus pattern C-C-x(2)-H-x(2)-C [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL known functional PA2's. However, this pattern will not detect some snake toxins homologous with PA2 but which have lost their catalytic activity as well as otoconin-22, a Xenopus protein from the aragonitic otoconia which is also unlikely to be enzymatically active. Other sequence(s) detected in SWISS-PROT 15. Consensus pattern [LIVMA]-C-{LIVMFYWPCST}-C-D-x(5)-C [D is the active site residue] Sequences known to belong to this class detected by the pattern the majority of functional and non-functional PA2's. Undetected sequences are bee PA2, gila monster PA2's, PA2 PL-X from habu and PA2 PA-5 from mulga. Other sequence(s) detected in SWISS-PROT 12. Expert(s) to contact by email Seilhamer J.J. jeff@incyte.com

Pfam	Prosite	Full Name	Description
			Last update November 1995 / Patterns and text revised. References [1] Davidson F.F., Dennis E.A. J. Mol. Evol. 31:228-238(1990). [2] Gomez F., Vandermeers A., Vandermeers-Piret M.-C., Herzog R., Rathe J., Stievenart M., Winand J., Christophe J. Eur. J. Biochem. 186:23-33(1989).
PI3_PI4_kinase	PDOC00710	Phosphatidylinositol 3- and 4-kinases signatures	Phosphatidylinositol 3-kinase (PI3-kinase) (EC 2.7.1.137) [1] is an enzyme that phosphorylates phosphoinositides on the 3-hydroxyl group of the inositol ring. The exact function of the three products of PI3-kinase - PI-3-P, PI-3,4-P(2) and PI-3,4,5-P(3) - is not yet known, although it is proposed that they function as second messengers in cell signalling. Currently, three forms of PI3-kinase are known: - The mammalian enzyme which is a heterodimer of a 110 Kd catalytic chain (p110) and an 85 Kd subunit (p85) which allows it to bind to activated tyrosine protein kinases. There are at least two different types of p100 subunits (alpha and beta). - Yeast TOR1/DRR1 and TOR2/DRR2 [2], PI3-kinases required for cell cycle activation. Both are proteins of about 280 Kd. - Yeast VPS34 [3], a PI3-kinase involved in vacuolar sorting and segregation. VPS34 is a protein of about 100 Kd. - Arabidopsis thaliana and soybean VPS34 homologs. Phosphatidylinositol 4-kinase (PI4-kinase) (EC 2.7.1.67) [4] is an enzyme that acts on phosphatidylinositol (PI) in the first committed step in the production of the second messenger inositol-1,4,5,-trisphosphate. Currently the following forms of PI4-kinases are known: - Human PI4-kinase alpha. - Yeast PIK1, a nuclear protein of 120 Kd. - Yeast STT4, a protein of 214 Kd. The PI3- and PI4-kinases share a well conserved domain at their C-terminal section; this domain seems to be distantly related to the catalytic domain of protein kinases [2]. We developed two signature patterns from the best conserved parts of this domain. Four additional proteins belong to this family: - Mammalian FKBP-rapamycin associated protein (FRAP) [5], which acts as the target for the cell-cycle arrest and immunosuppressive effects of the FKBP12-rapamycin complex. - Yeast protein ESR1 [6] which is required for cell growth, DNA repair and meiotic recombination. - Yeast protein TEL1 which is involved in controlling telomere length. - Yeast hypothetical protein YHR099w, a distantly related member of this family.

Pfam	Prosite	Full Name	Description
			- Fission yeast hypothetical protein SpAC22E12.16C. Description of pattern(s) and/or profile(s) Consensus pattern [LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q Sequences known to belong to this class detected by the pattern ALL, except for yeast YHR099w. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x- D-R-H-x(2)-N Sequences known to belong to this class detected by the pattern ALL, except for yeast YHR099w. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Hiles I.D., Otsu M., Volinia S., Fry M.J., Gout I., Dhand R., Panayotou G., Ruiz-Larrea F., Thompson A., Totty N.F., Hsuan J.J., Courtneidge S.A., Parker P.J., Waterfield M.D. Cell 70:419-429(1992). [2] Kunz J., Henriquez R., Schneider U., Deuter-Reinhard M., Movva N., Hall M.N. Cell 73:585-596(1993). [3] Schu P.V., Takegawa K., Fry M.J., Stack J.H., Waterfield M.D., Emr S.D. Science 260:88-91(1993). [4] Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N. EMBO J. 13:2352-2361(1994). [5] Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T., Lane W.S., Schreiber S.L. Nature 369:756-758(1994). [6] Kato R., Ogawa H. Nucleic Acids Res. 22:3104-3112(1994).
P-II	PDOC00439	P-II protein signatures	The P-II protein (gene glnB) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme. Conversely, in nitrogen excess, P-II-UMP is deuridylated and then promotes the adenylation of GS. P-II also indirectly controls the transcription of the GS gene (glnA) by preventing NR-II (ntrB) to phosphorylate NR-I (ntrC) which is the transcriptional activator of glnA. Once P-II is uridylylated, these events are reversed. P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is urydylated is located in the central part of the protein.

Pfam	Prosite	Full Name	Description
			In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being urydylated. In methanogenic archaebacteria, the nitrogenase iron protein gene (nifH) is followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation. In the red alga, Porphyra purpurea, there is a glnB homolog encoded in the chloroplast genome. Other proteins highly similar to glnB are: - Bacillus subtilis protein nrgB [5]. - Escherichia coli hypothetical protein ybal [6]. We developed two signature patterns for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the urydylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein. Description of pattern(s) and/or profile(s) Consensus pattern Y-[KR]-G-[AS]-[AE]-Y [The second Y is uridylated] Sequences known to belong to this class detected by the pattern ALL glnB's from eubacteria. Other sequence(s) detected in SWISS-PROT 4. Consensus pattern [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Magasanik B. Biochimie 71:1005-1012(1989). [2] Holtel A., Merrick M. Mol. Gen. Genet. 215:134-138(1988). [3] Cheah E., Carr P.D., Suffolk P.M., Vasuvedan S.G., Dixon N.E., Ollis D.L. Structure 2:981-990(1994). [4] Sibold L., Henriquet M., Possot O., Aubert J.-P. Res. Microbiol. 142:5-12(1991). [5] Wray L.V. Jr., Atkinson M.R., Fisher S.H. J. Bacteriol. 176:108-114(1994). [6] Allikmets R., Gerrard B.C., Court D., Dean M.C. Gene 136:231-236(1993).
pilin	PDOC00342	Prokaryotic N-terminal methylation site	A number of bacteria express filamentous adhesins known as pili. The pili are polar flexible filaments of about 5.4 nm diameter and 2500 nm average length;

Pfam	Prosite	Full Name	Description
			they consist of a single polypeptide chain (called pilin or fimbrial protein) arranged in a helical configuration of five subunits per turn in the assembled pilus. Gram-negative bacteria produce pilin which are characterized by the presence of a very short leader peptide of 6 to 7 residues, followed by a methylated N-terminal phenylalanine residue and by a highly conserved sequence of about 24 hydrophobic residues. This class of pilin is often referred to as NMePhe or type-4 pili [1,2]. Recently a number of bacterial proteins have been sequenced which share the following structural characteristics with type-4 pili [3]: a) The N-terminal residue, which is methylated, is hydrophobic (generally a phenylalanine or a methionine); b) The leader peptide is hydrophilic, consists of 5 to 10 residues (with two exceptions, see below) and ends with a glycine; c) The fifth residue of the mature sequence is a glutamate which seems to be required for the methylation step; d) The first twenty residues of the mature sequence are highly hydrophobic. These proteins are listed below: - Four proteins in an operon involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [4]. These proteins have been assigned a different gene name in each of the species where they have been sequenced: Species Gene names ----------------------- ------------------------ Aeromonas hydrophila exeG exeH exeI exeJ Erwinia chrysanthemi outG outH outI outJ Escherichia coli hofG hofH yheH yheI Klebsiella pneumoniae pulG pulH pulI pulJ Pseudomonase aeruginosa xcpT xcpU xcpV xcpW Vibrio cholerae epsG epsH epsI epsJ Xanthomonas campestris xpsG xpsH xpsI xpsJ - Vibrio cholerae toxin co-regulated pilin (gene tcpA). This pilin has a much longer putative leader peptide (25 residues). - Bacillus subtilis comG competence operon proteins 3, 4, and 5 which are involved for the uptake of DNA by competent Bacillus subtilis cells. - ppdA, ppdB and ppdC, three Escherichia coli hypothetical proteins found in the thyA-recC intergenic region. - ppdA, a hypothetical protein near the groeLS operon of Clostridium perfringens. The putative leader peptide is 23 residues long. We developed a signature pattern based on the N-terminal conserved region of all these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [KRHEQSTAG]-G-[FYLIVM]-[ST]-[LT]-[LIVP]-E-[LIVMFWSTAG](14) [The residue after the G is methylated]

Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / Text revised. References [1] Paranchych W., Frost L.S. Adv. Microb. Physiol. 29:53-114(1988). [2] Dalrymple B., Mattick J.S. J. Mol. Evol. 25:261-269(1987). [3] Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993). [4] Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).
PLA2_B		Lysophospholipase catalytic domain	Accession number: PF01735 Definition: Lysophospholipase catalytic domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_2127 (release 4.1) Gathering cutoffs: -283 -283 Trusted cutoffs: -185.70 -185.70 Noise cutoffs: -380.50 -380.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94299545 Reference Title: Delineation of two functionally distinct domains of Reference Title: cytosolic phospholipase A2, a regulatory Ca(2+)-dependent Reference Title: lipid-binding domain and a Ca(2+)-independent catalytic Reference Title: domain. Reference Author: Nalefski EA, Sultzman LA, Martin DM, Kriz RW, Towler PS, Reference Author: Knopf JL, Clark JD; Reference Location: J Biol Chem 1994;269:18239-18249. Reference Number: [2] Reference Medline: 94327513 Reference Title: The Saccharomyces cerevisiae PLB1 gene encodes a protein Reference Title: required for lysophospholipase and phospholipase B Reference Title: activity. Reference Author: Lee KS, Patton JL, Fido M, Hines LK, Kohlwein SD, Paltauf Reference Author: F, Henry SA, Levin DE; Reference Location: J Biol Chem 1994;269:19725-19730. Database Reference: SCOP; 1rlw; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002642; Database Reference PDB; 1bci ; 110; 138; Database Reference PDB; 1cjy B; 1110; 1430; Database Reference PDB; 1cjy A; 110; 498; Database Reference PDB; 1rlw ; 110; 140; Database Reference PDB; 1cjy B; 1463; 1497; Database Reference PDB; 1cjy B; 1539; 1717; Database Reference PDB; 1cjy A; 539; 721; Comment: This family consists of Lysophospholipase / phospholipase B Comment: EC:3.1.1.5 and cytosolic phospholipase A2 EC:3.1.4 which also Comment: has a C2 domain C2. Comment: Phospholipase B enzymes catalyse the release of fatty acids from

Pfam	Prosite	Full Name	Description
			Comment: lysophsopholipids and are capable in vitro of hydrolyzing all Comment: phospholipids extractable form yeast cells [1]. Comment: Cytosolic phospholipase A2 associates with natural membranes in Comment: response to physiological increases in Ca2+ and selectively Comment: hydrolyses arachidonyl phospholipids [2], the aligned region Comment: corresponds the the carboxy-terminal Ca2+-independent catalytic Comment: domain of the protein as discussed in [2]. Number of members: 23
PLAT		PLAT/LH2 domain	Accession number: PF01477 Definition: PLAT/LH2 domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 25 25 Trusted cutoffs: 29.40 29.40 Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference: SCOP; 1lpa; fa; [SCOP-USA][CATH-PDBSUM] Database reference: PROSITE_PROFILE; PS50095; Database Reference INTERPRO; IPR001024; Database Reference PDB; 1lox ; 2; 112; Database Reference PDB; 1hpl B; 336; 445; Database Reference PDB; 1hpl A; 338; 447; Database Reference PDB; 1eth C; 337; 403; Database Reference PDB; 1eth A; 339; 405; Database Reference PDB; 1eth C; 403; 445; Database Reference PDB; 1eth A; 405; 447; Database Reference PDB; 1rp1 ; 339; 449; Database Reference PDB; 1bu8 A; 340; 407; Database Reference PDB; 1bu8 A; 415; 452; Database Reference PDB; 1gpl ; 322; 334; Database Reference PDB; 1ca1 ; 256; 370; Database Reference PDB; 1qm6 A; 256; 370; Database Reference PDB; 1qm6 B; 256; 370; Database Reference PDB; 1qmd A; 256; 370; Database Reference PDB; 1qmd B; 256; 370; Comment: This domain is found in a variety of membrane or Comment: lipid associated proteins. It is called the PLAT Comment: (Polycystin-1, Lipoxygenase, Alpha-Toxin) domain or Comment: LH2 (Lipoxygenase homology) domain. The known structure Comment: of pancreatic lipase shows this domain binds to procolipase Comment: Colipase, which mediates membrane association. Comment: So it appears possible that this domain mediates membrane Comment: attachment via other protein binding partners. The Comment: structure of this domain is known for many members of the Comment: family and is composed of a beta sandwich. Number of members: 82
PLRV_ORF5		Potato leaf roll virus readthrough protein	Accession number: PF01690 Definition: Potato leaf roll virus readthrough protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1335 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 116.40 116.40 Noise cutoffs: -285.50 -285.50

Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94233771 Reference Title: Changes in the amino acid sequence of the coat protein Reference Title: readthrough domain of potato leafroll luteovirus affect the Reference Title: formation of an epitope and aphid transmission. Reference Author: Jolly CA, Mayo MA; Reference Location: Virology 1994;201:182-185. Database Reference INTERPRO; IPR002929; Comment: This family consists mainly of the potato leaf roll virus Comment: readthrough protein. This is generated via a readthrough Comment: of open reading frame 3 a coat protein allowing transcription Comment: of open reading frame 5 to give an extended coat protein Comment: with a large c-terminal addition or read through domain [1]. Comment: The readthrough protein is thought to play a role in the Comment: circulative aphid transmission of potato leaf roll virus [1]. Comment: Also in the family is open reading frame 6 from beet western Comment: yellows virus and potato leaf roll virus both luteovirus and Comment: an unknown protein from cucurbit aphid-borne yellows virus a Comment: closterovirus. Number of members: 28
PMSR		Peptide methionine sulfoxide reductase	Accession number: PF01625 Definition: Peptide methionine sulfoxide reductase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1111 (release 4.1) Gathering cutoffs: -62 -62 Trusted cutoffs: -28.00 -28.00 Noise cutoffs: -96.70 -96.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96353931 Reference Title: Peptide methionine sulfoxide reductase contributes to the Reference Title: maintenance of adhesins in three major pathogens. Reference Author: Wizemann TM, Moskovitz J, Pearce BJ, Cundell D, Arvidson Reference Author: CG, So M, Weissbach H, Brot N, Masure HR; Reference Location: Proc Natl Acad Sci USA 1996;93:7985-7990. Reference Number: [2] Reference Medline: 96312545 Reference Title: Cloning the expression of a mammalian gene involved in the Reference Title: reduction of methionine sulfoxide residues in proteins. Reference Author: Moskovitz J, Weissbach H, Brot N; Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-2099. Database Reference INTERPRO; IPR002569; Comment: This enzyme repairs damaged proteins. Methionine sulfoxide in proteins Comment: is reduced to methionine. Number of members: 28
Pollen_allerg_2		Ribonuclease (pollen	Accession number: PF01620

Pfam	Prosite	Full Name	Description
		allergen)	Definition: Ribonuclease (pollen allergen) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1050 (release 4.1) Gathering cutoffs: -3 -3 Trusted cutoffs: 23.10 23.10 Noise cutoffs: -29.40 -29.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95246885 Reference Title: Major allergen Phl p Vb in timothy grass is a novel pollen Reference Title: RNase. Reference Author: Bufe A, Schramm G, Keown MB, Schlaak M, Becker WM; Reference Location: Febs lett 1995;363:6-12. Database Reference INTERPRO; IPR002914; Database reference: PFAMB; PB037130; Comment: This family contains grass pollen proteins of group V. Comment: Swiss:Q40963 has been shown to possess ribonuclease Comment: activity [1]. Number of members: 27
POR_N		Pyruvate flavodoxin/ferredoxin oxidoreductase (N terminus)	Accession number: PF01855 Definition: Pyruvate flavodoxin/ferredoxin oxidoreductase (N terminus) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_323 (release 4.2) Gathering cutoffs: -116 -116 Trusted cutoffs: -113.60 -113.60 Noise cutoffs: -119.50 -119.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96125254 Reference Title: Molecular and phylogenetic characterization of pyruvate and Reference Title: 2-ketoisovalerate ferredoxin oxidoreductases from Reference Title: Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase Reference Title: from Thermotoga maritima. Reference Author: Kletzin A, Adams MW; Reference Location: J Bacteriol 1996;178:248-257. Reference Number: [2] Reference Medline: 94022264 Reference Title: Growth of the cyanobacterium Anabaena on molecular Reference Title: nitrogen: NifJ is required when iron is limited. Reference Author: Bauer CC, Scappino L, Haselkorn R; Reference Location: Proc Natl Acad Sci U S A 1993;90:8812-8816. Reference Number: [3] Reference Medline: 99140300 Reference Title: Crystal structures of the key anaerobic enzyme Reference Title: pyruvate:ferredoxin oxidoreductase, free and in complex Reference Title: with pyruvate. Reference Author: Chabriere E, Charon MH, Volbeda A, Pieulle L, Hatchikian Reference Author: EC, Fontecilla-Camps JC; Reference Location: Nat Struct Biol 1999;6:182-190. Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002880; Database Reference PDB; 1b0p A; 43; 328;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1b0p B; 43; 328; Database Reference PDB; 2pda A; 43; 328; Database Reference PDB; 2pda B; 43; 328; Database reference: PFAMB; PB014847; Comment: This family includes the N terminal region of the pyruvate ferredoxin Comment: oxidoreductase, corresponding to the first two structural domains. Comment: This region is involved in inter subunit contacts [3]. Pyruvate Comment: oxidoreductase (POR) catalyses the final step in the fermentation Comment: of carbohydrates in anaerobic microorganisms [1]. This involves the Comment: oxidative decarboxylation of pyruvate with the participation of Comment: thiamine followed by the transfer of an acetyl moiety to coenzyme Comment: A for the synthesis of acetyl-CoA [1]. The family also includes Comment: pyruvate flavodoxin oxidoreductase as encoded by the nifJ gene in Comment: cyanobacterium which is required for growth on molecular nitrogen Comment: when iron is limited [2]. Number of members: 55
PPE		PPE family	Accession number: PF00823 Definition: PPE family Author: Bateman A Alignment method of seed: Clustalw_manual Source of seed members: Pfam-B_297 (release 3.0) Gathering cutoffs: -90 -90 Trusted cutoffs: -88.20 -88.20 Noise cutoffs: -105.30 -105.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98295987 Reference Title: Deciphering the biology of Mycobacterium tuberculosis from Reference Title: the complete genome sequence. Reference Author: Reference Location: Nature 1998;393:537-544. Database Reference INTERPRO; IPR000030; Database reference: PFAMB; PB040834; Comment: This family named after a PPE motif near to the amino Comment: terminus of the domain. The PPE family of proteins Comment: all contain an amino-terminal region of about 180 Comment: amino acids. The carboxyl terminus of this family Comment: are variable, and on the basis of this region fall Comment: into at least three groups. The MPTR subgroup has Comment: tandem copies of a motif NXGXGNXG. The second subgroup Comment: contains a conserved motif at about position 350. Comment: The third group are only related in the amino terminal Comment: region. Comment: The function of these proteins is uncertain but it Comment: has been suggested that they may be related to Comment: antigenic variation of Mycobacterium tuberculosis [1]. Number of members: 75
PRA-CH		Phosphoribosyl-AMP	Accession number: PF01502

Pfam	Prosite	Full Name	Description
		cyclohydrolase	Definition: Phosphoribosyl-AMP cyclohydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_782 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 88.20 88.20 Noise cutoffs: -44.30 -44.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99129952 Reference Title: N1-(5'-phosphoribosyl)adenosine-5'-monophosphate Reference Title: cyclohydrolase: purification and characterization of a Reference Title: unique metalloenzyme. Reference Author: D'Ordine RL, Klem TJ, Davisson VJ; Reference Location: Biochemistry 1999;38:1537-1546. Database Reference INTERPRO; IPR002496; Comment: This enzyme catalyses the third step in the histidine Comment: biosynthetic pathway. It requires Zn ions for activity. Number of members: 28
PRA-PH		Phosphoribosyl-ATP pyrophosphohydrolase	Accession number: PF01503 Definition: Phosphoribosyl-ATP pyrophosphohydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_784 (release 4.0) Gathering cutoffs: 6 6 Trusted cutoffs: 12.10 12.10 Noise cutoffs: 1.00 1.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 79216449 Reference Title: The product of the his4 gene cluster in Saccharomyces Reference Title: cerevisiae. A trifunctional polypeptide. Reference Author: Keesey JK Jr, Bigelis R, Fink GR; Reference Location: J Biol Chem 1979 Aug 10;254:7427-7433. Reference Number: [2] Reference Medline: 86310274 Reference Title: Primary and secondary structural homologies between the Reference Title: HIS4 gene product of Saccharomyces cerevisiae and the hisIE Reference Title: and hisD gene products of Escherichia coli and Salmonella Reference Title: typhimurium. Reference Author: Bruni CB, Carlomagno MS, Formisano S, Paolella G; Reference Location: Mol Gen Genet 1986;203:389-396. Database Reference INTERPRO; IPR002497; Comment: This enzyme catalyses the second step in the histidine Comment: biosynthetic pathway. Number of members: 32
PseudoU_synth_1		tRNA pseudouridine synthase	Accession number: PF01416 Definition: tRNA pseudouridine synthase Previous Pfam IDs: PseudoU_synt; Author: Howe K Alignment method of seed: Clustalw Source of seed members: swissprot Gathering cutoffs: 30 30 Trusted cutoffs: 39.10 39.10 Noise cutoffs: -55.00 -55.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98254513

Pfam	Prosite	Full Name	Description
			Reference Title: Transfer RNA-pseudouridine synthetase Pus1 of Saccaromyces Reference Title: cerevisiae contains one atom of zinc essential for its Reference Title: native conformation and tRNA recognition. Reference Author: Arluison V, Hountondji C, Robert B, Grosjean H; Reference Location: Biochemistry 1998;37:7268-7276. Database Reference INTERPRO; IPR001406; Database reference: PFAMB; PB027500; Comment: Involved in the formation of pseudouridine at the anticodon stem Comment: and loop of transfer-RNAs Comment: Pseudouridine is an isomer of uridine (5-(beta-D-ribofuranosyl) Comment: uracil, and id the most abundant modified nucleoside found in Comment: all cellular RNAs. Comment: The TruA-like proteins also exhibit a conserved sequence with Comment: a strictly conserved aspartic acid, likely involved in catalysis Number of members: 31
PseudoU_synth_2		RNA pseudouridylate synthase	Accession number: PF00849 Definition: RNA pseudouridylate synthase Previous Pfam IDs: YABO; Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_421 (release 3.0) Gathering cutoffs: 20 20 Trusted cutoffs: 20.90 20.90 Noise cutoffs: -44.40 -44.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96079974 Reference Title: A dual-specificity pseudouridine synthase: an Escherichia Reference Title: coli synthase purified and cloned on the basis of its Reference Title: specificity for psi 746 in 23S RNA is also specific for psi Reference Title: 32 in tRNA(phe). Reference Author: Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J; Reference Location: RNA 1995;1:437-448. Database Reference: PROSITE; PDOC00869 Database Reference: PROSITE; PDOC00885 Database Reference INTERPRO; IPR000613; Database reference: PFAMB; PB041160; Database reference: PFAMB; PB041232; Comment: Members of this family are involved in modifying bases in RNA molecules. Comment: They carry out the conversion of uracil bases to pseudouridine. This family Comment: includes RluD Swiss:P33643, a pseudouridylate synthase that converts Comment: specific uracils to pseudouridine in 23S rRNA. RluA from E. coli Comment: converts bases in both rRNA and tRNA [1]. Number of members: 78
PWI		PWI domain	Accession number: PF01480 Definition: PWI domain Author: Bateman A Alignment method of seed: Clustalw_manual Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 64.40 64.40 Noise cutoffs: -3.50 -3.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1]

Pfam	Prosite	Full Name	Description
			Reference Medline: 10322432 Reference Title: The PWI motif: a new protein domain in splicing factors. Reference Author: Blencowe BJ, Ouzounis CA; Reference Location: Trends Biochem Sci 1999;24:179-180. Database Reference INTERPRO; IPR002483; Number of members: 11
R3H		R3H domain	Accession number: PF01424 Definition: R3H domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Medline:99003905 Gathering cutoffs: 25 25 Trusted cutoffs: 59.30 59.30 Noise cutoffs: 5.10 5.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99003905 Reference Title: The R3H motif: a domain that binds single-stranded nucleic Reference Title: acids. Reference Author: Grishin NV; Reference Location: Trends Biochem Sci 1998;23:329-330. Database Reference INTERPRO; IPR001374; Database reference: PFAMB; PB041444; Comment: The name of the R3H domain comes from the characteristic spacing Comment: of the most conserved arginine and histidine residues. The Comment: function of the domain is predicted to be binding ssDNA. Number of members: 28
RepB_protein		Initiator RepB protein	Accession number: PF01051 Definition: Initiator RepB protein Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members. Pfam-B_313 (release 3.0) Gathering cutoffs: 14 14 Trusted cutoffs: 19.00 16.20 Noise cutoffs: 11.80 12.90 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98284148 Reference Title: Replication and control of circular bacterial plasmids. Reference Author: del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Reference Author: Diaz-Orejas R; Reference Location: Microbiol Mol Biol Rev 1998;62:434-464. Reference Number: [2] Reference Medline: 97324207 Reference Title: Initiation of replication of plasmid pMV158: mechanisms of Reference Title: DNA strand-transfer reactions mediated by the initiator Reference Title: RepB protein. Reference Author: Moscoso M, Eritja R, Espinosa M; Reference Location: J Mol Biol 1997;268:840-856. Database Reference INTERPRO; IPR000525; Database Reference PDB; 1rep C; 198; 240; Database reference: PFAMB; PB000509; Comment: This protein is an initiator of plasmid replication. Comment: RepB possesses nicking-closing (topoisomerase I) like activity. Comment: It is also able to perform a strand transfer reaction on ssDNA Comment: that contains its target. Number of members: 51

Pfam	Prosite	Full Name	Description
Rhomboid		Rhomboid family	Accession number: PF01694 Definition: Rhomboid family Author: Sohrmann M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B 1399 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 143.60 143.60 Noise cutoffs: -43.60 -43.60 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 90249726 Reference Title: rhomboid, a gene required for dorsoventral axis Reference Title: establishment and peripheral nervous system development in Reference Title: Drosophila melanogaster. Reference Author: Bier E, Jan LY, Jan YN; Reference Location: Genes Dev 1990;4:190-203. Database Reference INTERPRO; IPR002610; Database reference: PFAMB; PB041113; Comment: This family contains integral membrane proteins that are Comment: related to Drosophila rhomboid protein Swiss:P20350. Members Comment: of this family are found in bacteria and eukaryotes. These Comment: proteins contain three strongly conserved histidines in the Comment: putative transmembrane regions that may be involved in the Comment: as yet unknown function of these proteins. Number of members: 27
Ribosomal_L18ae		Ribosomal L18ae protein family	Accession number: PF01775 Definition: Ribosomal L18ae protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST Q02543 Gathering cutoffs: 25 25 Trusted cutoffs: 136.70 136.70 Noise cutoffs: -99.80 -99.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002670; Number of members: 11
Ribosomal_L21p	PDOC00899	Ribosomal protein L21 signature	Ribosomal protein L21 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L21 is known to bind to the 23S rRNA in the presence of L20. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L21. - Marchantia polymorpha chloroplast L21. - Cyanelle L21. - Spinach chloroplast L21 (nuclear-encoded). Eubacterial L21 is a protein of about 100 amino-acid residues, the mature form of the spinach chloroplast L21 has 200 residues. As a signature pattern, we selected a conserved region located in the C-terminal section of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-[RQ]-x(2)-[ST]

Pfam	Prosite	Full Name	Description
			Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised.
Ribosomal_L22e		Ribosomal L22e protein family	Accession number: PF01776 Definition: Ribosomal L22e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P56628 Gathering cutoffs: 25 25 Trusted cutoffs: 262.80 262.80 Noise cutoffs: -52.00 -52.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002671; Number of members: 11
Ribosomal_L27e		Ribosomal L27e protein family	Accession number: PF01777 Definition: Ribosomal L27e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P51419 Gathering cutoffs: 25 25 Trusted cutoffs: 326.90 326.90 Noise cutoffs: -47.80 -47.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR001141; Number of members: 9
Ribosomal_L29	PDOC00501	Ribosomal protein L29 signature	Ribosomal protein L29 is one of the proteins from the large ribosomal subunit. L29 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L29. - Red algal L29. - Archaebacterial L29. - Mammalian L35 - Caenorhabditis elegans L35 (ZK652.4). - Yeast L35. L29 is a protein of 63 to 138 amino-acid residues. As a signature pattern, we selected a conserved region located in the central section of L29. Description of pattern(s) and/or profile(s) Consensus pattern [KNQS]-[PSTLN]-x(2)-[LIMFA]-[KRGSAN]-x-[LIVYSTA]-[KR]- [KRHQS]-[DESTANRL]-[LIV]-A-[KRCQVT]-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_L31e	PDOC00881	Ribosomal protein L31e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L31 [1]. - Chlamydomonas reinhardtii L31. - Yeast L34.

Pfam	Prosite	Full Name	Description
			- Halobacterium marismortui HL30 [2]. These proteins have 87 to 128 amino-acid residues. As a signature pattern, we selected a conserved region located in the central section. Description of pattern(s) and/or profile(s) Consensus pattern V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AKH]-x-W-x-[KR]-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Tanaka T., Kuwano Y., Kuzumaki T., Ishikawa K., Ogata K. Eur. J. Biochem. 162:45-48(1987). [2] Bergmann U., Arndt E. Biochim. Biophys. Acta 1050:56-60(1990).
Ribosomal_L35Ae	PDOC00849	Ribosomal protein L35Ae signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Vertebrate L35A. - Caenorhabditis elegans L35A (F10E7.7). - Yeast L37A/L37B (Rp47). - Pyrococcus woesei L35A homolog [1]. These proteins have 87 to 110 amino-acid residues. As a signature pattern, we selected a highly conserved stretch of 22 residues in the C-terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Ouzounis C., Kyrpides N., Sander C. Nucleic Acids Res. 23:565-570(1995).
Ribosomal_L35p	PDOC00721	Ribosomal protein L35 signature	Ribosomal protein L35 is one of the proteins from the large subunit of the ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L35. - Plant chloroplast L35 (nuclear-encoded). - Red algal chloroplast L35. - Cyanelle L35. L35 is a basic protein of 60 to 70 amino-acid residues. As a signature pattern we selected a conserved region in the N-terminal section.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-K-[TV]-x(2)-[GSA]-[SAILV]-x-K-R-[LIVMFY]-[KRLS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_L36e	PDOC00916	Ribosomal protein L36e signature	A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L36 [1]. - Drosophila L36 (M(1)1B). - Caenorhabditis elegans L36 (F37C12.4). - Candida albicans L39. - Yeast YL39. These proteins have 99 to 104 amino acids. As a signature pattern, we selected a conserved region in the central part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / First entry. References [1] Chan Y.-L., Paz V., Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 192:849-853(1993).
Ribosomal_L37ae		Ribosomal L37ae protein family	Accession number: PF01780 Definition: Ribosomal L37ae protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P54051 Gathering cutoffs: 25 25 Trusted cutoffs: 145.10 145.10 Noise cutoffs: -46.90 -46.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002674; Comment: This ribosomal protein is found in archaebacteria and Comment: eukaryotes. It contains four conserved cysteine Comment: residues that may bind to zinc. Number of members: 15
Ribosomal_L37e	PDOC00827	Ribosomal protein L37e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L37 [1]. - Leishmania infantum L37 [2]. - Fission yeast YL35 [3]. - Halobacterium marismortui L37e (L35e) [4]. These proteins have 56 to 96 amino-acid residues. As a signature pattern, we

Pfam	Prosite	Full Name	Description
			selected a highly conserved region located in the N-terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern G-T-x-[SA]-x-G-x-[KR]-x(3)-[STLR]-x(0,1)-H-x(2)-C-x-R-C-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Chan Y.-L., Paz V., Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 192:590-596(1993). [2] Myler P.J., Tripp C.A., Thomas L., Venkataraman G.M., Merlin G., Stuart K. Mol. Biochem. Parasitol. 62:147-152(1993). [3] Otaka E., Higo K.-I., Itoh T. Mol. Gen. Genet. 191:519-524(1983). [4] Bergmann U., Wittmann-Liebold B. Biochim. Biophys. Acta 1173:195-200(1993).
Ribosomal_L38e		Ribosomal L38e protein family	Accession number: PF01781 Definition: Ribosomal L38e protein family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P23411 Gathering cutoffs: 25 25 Trusted cutoffs: 127.60 127.60 Noise cutoffs: -24.50 -24.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91207349 Reference Title: The primary structure of rat ribosomal protein L38. Reference Author: Kuwano Y, Olvera J, Wool IG; Reference Location: Biochem Biophys Res Commun 1991;175:551-555. Database Reference INTERPRO; IPR002675; Number of members: 8
Ribosomal_L39	PDOC00050	Ribosomal protein L39e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L39 [1]. - Plants L39. - Yeast L46 [2]. - Archebacterial L39e [3]. These proteins are very basic. About 50 residues long, they are the smallest proteins of eukaryotic-type ribosomes. As a signature pattern, we selected a conserved region in the C-terminal section of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-

Pfam	Prosite	Full Name	Description
			[NHY]-W-R-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Pattern and text revised. References [1] Lin A., McNally J., Wool I.G. J. Biol. Chem. 259:487-490(1984). [2] Leer R.J., van Raamsdonk-Duin M.M.C., Kraakman P., Mager W.H., Planta R.J. Nucleic Acids Res. 13:701-709(1985). [3] Ramirez C., Louie K.A., Matheson A.T. FEBS Lett. 250:416-418(1989).
Ribosomal_L4	PDOC00724	Ribosomal protein L1e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists [1,2,3, 4] of: - Vertebrate L1 (L4). - Drosophila L1. - Plant L1. - Yeast L2 (Rp2). - Fission yeast L2. - Halobacterium marismortui HmaL4 (HL6). - Methanococcus jannaschii MJ0177. These proteins have 246 (archaebacteria) to 427 (human) amino acids. As a signature pattern, we selected a conserved region in the N-terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern N-x(3)-[KRM]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-x(7)-[RK]-[GS]-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Rafti F., Gargiulo G., Manzi A., Malva C., Graziani F. Nucleic Acids Res. 17:456-456(1989). [2] Presutti C., Villa T., Bozzoni I. Nucleic Acids Res. 21:3900-3900(1993). [3] Bagni C., Mariottini P., Annesi F., Amaldi F. Arndt E., Kroemer W., Hatakeyama T. Biochim. Biophys. Acta 1216:475-478(1993). J. Biol. Chem. 265:3034-3039(1990).
Ribosomal_S20p		Ribosomal protein S20	Accession number: PF01649 Definition: Ribosomal protein S20 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1685 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 57.30 57.30 Noise cutoffs: -25.50 -25.50

Pfam	Prosite	Full Name	Description
			HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 88230452 Reference Title: Interaction of proteins S16, S17 and S20 with 16 S Reference Title: ribosomal RNA. Reference Author: Stern S, Changchien LM, Craven GR, Noller HF; Reference Location: J Mol Biol 1988;200:291-299. Database Reference INTERPRO; IPR002583; Comment: Bacterial ribosomal protein S20 interacts with 16S rRNA [1]. Number of members: 29
Ribosomal_S27e	PDOC00898	Ribosomal protein S27e signature	A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of [1]: - Mammalian S27 (human S27 was originally known as metallopan-stimulin 1). - Chlamydomonas reinhardtii S27. - Entamoeba histolytica S27. - Yeast S27. - Archaebacterial S27e. These proteins have from 62 to 87 amino acids. They contain, in their central section, a putative zinc-finger region of the type C-x(2)-C-x(14)-C-x(2)-C. We have selected that region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [QKT]-C-x(2)-C-x(6)-F-[GSD]-x-[PSA]-x(5)-C-x(2)-C-[GSA]- x(2)-[LV]-x(2)-P-x-G [The four C's are potential zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Chan Y.-L., Suzuki K., Olvera J., Wool I.G. Nucleic Acids Res. 21:649-655(1993).
Ribosomal_S3_C	PDOC00474	Ribosomal protein S3 signature	Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S3. - Algal and plant chloroplast S3. - Cyanelle S3. - Archaebacterial S3. - Plant mitochondrial S3. - Vertebrate S3. - Insect S3. - Caenorhabditis elegans S3 (C23G10.3). - Yeast S3 (Rp13). S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]- x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Hallick R.B. hallick@arizona.edu Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
Ribosomal_S3_N	PDOC00474	Ribosomal protein S3 signature	Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S3. - Algal and plant chloroplast S3. - Cyanelle S3. - Archaebacterial S3. - Plant mitochondrial S3. - Vertebrate S3. - Insect S3. - Caenorhabditis elegans S3 (C23G10.3). - Yeast S3 (Rp13). S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section. Description of pattern(s) and/or profile(s) Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]- x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Hallick R.B. hallick@arizona.edu Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).
RimM		RimM	Accession number: PF01782 Definition: RimM Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P51419 Gathering cutoffs: 25 25 Trusted cutoffs: 49.00 49.00 Noise cutoffs: -66.10 -66.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98083058 Reference Title: RimM and RbfA are essential for efficient processing of 16S Reference Title: rRNA in Escherichia coli. Reference Author: Bylund GO, Wipemo LC, Lundberg LA,

Pfam	Prosite	Full Name	Description
			Wikstrom PM; Reference Location: J Bacteriol 1998;180:73-82. Database Reference INTERPRO; IPR002676; Comment: The RimM protein is essential for efficient processing of 16S rRNA [1]. Comment: The RimM protein was shown to have affinity for free ribosomal 30S Comment: subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14
RNA_dep_RNA_pol		RNA dependent RNA polymerase	Accession number: PF00680 Definition: RNA dependent RNA polymerase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_32 (release 2.1) Gathering cutoffs: -127 -127 Trusted cutoffs: -117.00 -117.00 Noise cutoffs: -137.30 -137.30 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference: SCOP; 1rdr; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001205; Database Reference PDB; 1rdr ; 12; 37; Database Reference PDB; 1rdr ; 182; 460; Database Reference PDB; 1rdr ; 67; 97; Database reference: PFAMB; PB039844; Database reference: PFAMB; PB040630; Database reference: PFAMB; PB040631; Database reference: PFAMB; PB040844; Database reference: PFAMB; PB041022; Database reference: PFAMB; PB041498; Number of members: 271
RNA_dep_RNApol2		RNA dependent RNA polymerase	Accession number: PF00978 Definition: RNA dependent RNA polymerase Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_13 (release 3.0) Gathering cutoffs: 8.5 0 Trusted cutoffs: 8.50 0.20 Noise cutoffs: 8.40 8.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93188140 Reference Title: Roles of nonstructural polyproteins and cleavage products Reference Title: in regulating Sindbis virus RNA replication and Reference Title: transcription. Reference Author: Lemm JA, Rice CM; Reference Location: J Virol 1993;67:1916-1926. Reference Number: [2] Reference Medline: 96323143 Reference Title: Complete replication in vitro of tobacco mosaic virus RNA Reference Title. by a template-dependent, membrane-bound RNA polymerase. Reference Author: Osman TA, Buck KW; Reference Location: J Virol 1996;70:6227-6234. Reference Number: [3] Reference Medline: 94047331 Reference Title: Bromovirus RNA replication and transcription require Reference Title: compatibility between the polymerase- and helicase-like Reference Title: viral RNA synthesis proteins. Reference Author: Dinant S, Janda M, Kroner PA, Ahlquist P; Reference Location: J Virol 1993;67:7181-7189. Reference Number: [4] Reference Medline: 94094568 Reference Title: Evolution and taxonomy of positive-strand

Pfam	Prosite	Full Name	Description
			RNA viruses: Reference Title: implications of comparative analysis of amino acid Reference Title: sequences. Reference Author: Koonin EV, Dolja VV; Reference Location: Crit Rev Biochem Mol Biol 1993;28:375-430. Database Reference INTERPRO; IPR001788; Database reference: PFAMB; PB000096; Database reference: PFAMB; PB006751; Comment: This family may represent an RNA dependent RNA polymerase. Comment: The family contains the following proteins: Comment: 2A protein from bromoviruses Comment: putative RNA dependent RNA polymerase from tobamoviruses Comment: Non structural polyprotein from togaviruses Number of members: 125
RNA_pol	PDOC00410	Bacteriophage-type RNA polymerase family active site signatures	Many forms of RNA polymerase (EC 2.7.7.6) are known. Most RNA polymerases are multimeric enzymes, but there is a family of single chain polymerases, which are evolutionary related, and which originate from bacteriophages or from mitochondria. The RNA polymerases that belong to this family are [1]: - Podoviridae bacteriophages T3, T7, and K11 polymerase. - Bacteriophage SP6 polymerase. - Vertebrate mitochondrial polymerase (gene POLRMT). - Fungal mitochondrial polymerase (gene RPO41). - Polymerases encoded on mitochondrial linear DNA plasmids in various fungi and plants: Agaricus bitorquis pEM, Claviceps purpurea pClK1, Neurospora crassa Kalilo; Neurospora intermedia Maranhar and maize S-2). Two conserved aspartate and one lysine residue have been shown [2,3] to be part of the active site of T7 polymerase. We have used the regions around the first aspartate and around the lysine as signature patterns for this family of polymerases. Description of pattern(s) and/or profile(s) Consensus pattern P-[LIVM]-x(2)-D-[GA]-[ST]-[AC]-[SN]-[GA]-[LIVMFY]-Q [D is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVMF]-x-R-x(3)-K-x(2)-[LIVMF]-M-[PT]-x(2)-Y [K is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Text revised. References [1] McAllister W.T., Raskin C.A. Mol. Microbiol. 10:1-6(1993). [2] Maksimova T.G., Mustayev A.A., Zaychikov E.F., Lyakhov D.L., Tunitskaya V.L., Akbarov A.K., Luchin S.V., Rechinsky V.O., Chernov B.K., Kochetkov S.N. Eur. J. Biochem. 195:841-847(1991).

Pfam	Prosite	Full Name	Description
			[3] Sousa R., Chung Y.J., Rose J.P., Wang B.-C. Nature 364:593-599(1993).
RNA_pol_A		RNA polymerase alpha subunit	Accession number: PF00623 Definition: RNA polymerase alpha subunit Author: Bateman A Alignment method of seed: HMM_built_from_alignment Source of seed members: Pfam-B_3 (release 2.1) Gathering cutoffs: 9 0 Trusted cutoffs: 13.50 2.90 Noise cutoffs: 8.50 8.50 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97066998 Reference Title: Structural modules of the large subunits of RNA polymerase. Reference Title: Introducing archaebacterial and chloroplast split sites in Reference Title: the beta and beta' subunits of Escherichia coli RNA Reference Title: polymerase. Reference Author: Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst Reference Author: SA, Goldfarb A; Reference Location: J Biol Chem 1996;271:27969-27974. Database Reference INTERPRO; IPR000722; Database reference: PFAMB; PB003218; Comment: -!- RNA polymerases catalyse the DNA dependent polymerisation Comment: of RNA. Prokaryotes contain a single RNA polymerase Comment: compared to three in eukaryotes (not including mitochondrial. Comment: and chloroplast polymerases). Comment: -!- Members of this family include: Comment: A subunit from eukaryotes Comment: gamma subunit from cyanobacteria Comment: beta' subunit from eubacteria Comment: A' subunit from archaebacteria Comment: B" from chloroplasts Number of members: 202
RNA_pol_A2		RNA polymerase A/beta'/A" subunit	Accession number: PF01854 Definition: RNA polymerase A/beta'/A" subunit Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_288 (release 4.2) Gathering cutoffs: -120 -120 Trusted cutoffs: -116.50 -116.50 Noise cutoffs: -125.00 -125.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 88335550 Reference Title: Relatedness of archaebacterial RNA polymerase core subunits Reference Title: to their eubacterial and eukaryotic equivalents. Reference Author: Berghofer B, Krockel L, Kortner C, Truss M, Schallenberg J, Reference Author: Klein A; Reference Location: Nucleic Acids Res 1988;16:8113-8128. Database Reference INTERPRO; IPR002879; Database reference: PFAMB; PB000546; Database reference: PFAMB; PB000846; Database reference: PFAMB; PB000984; Database reference: PFAMB; PB001168; Comment: RNA polymerases catalyse the DNA dependent polymerisation Comment: of RNA. Prokaryotes contain a single RNA polymerase Comment: compared to three in eukaryotes (not

Pfam	Prosite	Full Name	Description
			including mitochondrial. Comment: and chloroplast polymerases). Comment: This family includes a region of about 400 amino acids. Comment: This family includes the whole archaebacterial A" subunit, Comment: but only the C terminal region of the A subunit from eukaryotes Comment: and the beta' subunit from eubacteria. Number of members: 105
RNB	PDOC00904	Ribonuclease II family signature	On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family: - Escherichia coli and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene rnb) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction. - Bacterial ribnuclease R [2], a 3'-5'exoribonuclease that participates in an essential cell function. - Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase. - Yeast protein DIS3 [3], which binds to ran (GSP1) and ehances the the nucleotide-releasing activity of RCC1 on ran. - Fission yeast protein dis3, which is implicated in mitotic control. - Neurospora crassa cyt-4, a mitochondrial protein required for RNA 5' and 3' end processing and splicing. - Yeast protein MSU1, which is involved in mitochondrial biogenesis. - Synechocystis strain PCC 6803 protein zam [4], which control resistance to the carbonic anhydrase inhibitor acetazolamide. - Caenorhabditis elegans hypothetical protein F48E8.6. The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C- terminal section [5]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. We have developed a signature pattern based on the core of this conserved domain. Description of pattern(s) and/or profile(s) Consensus pattern [HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STALV]-x-[FWVAC]-[TV]- [SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Zilhao R., Camelo L., Arraiano C.M. Mol. Microbiol. 8:43-51 (1993). [2] Cheng Z.-F., Zuo Y., Li Z., Rudd K.E., Deutscher M P. J. Biol. Chem. 273:14077-14080(1998). [3]

Pfam	Prosite	Full Name	Description
			Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M., Nakashima N., Yanagida M., He X., Mueller U., Sazer S., Nishimoto T. EMBO J. 15:5595-5605(1996). [4] Beuf L., Bedu S., Cami B., Joset F. Plant Mol. Biol. 27:779-788(1995). [5] Mian I.S. Nucleic Acids Res. 25:3187-3195(1997).
RRF		Ribosome recycling factor	Accession number: PF01765 Definition: Ribosome recycling factor Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_949 (release 4.2) Gathering cutoffs: -35 -35 Trusted cutoffs: -34.90 -34.90 Noise cutoffs: -76.20 -76.20 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94240115 Reference Title: Ribosome recycling factor (ribosome releasing factor) is Reference Title: essential for bacterial growth. Reference Author: Janosi L, Shimizu I, Kaji A; Reference Location: Proc Natl Acad Sci U S A 1994;91:4249-4253. Database Reference INTERPRO; IPR002661; Comment: The ribosome recycling factor (RRF / ribosome release factor) dissociates Comment: the ribosome from the mRNA after termination of translation, and is Comment: essential bacterial growth [1]. Thus ribosomes are "recycled" and ready Comment: for another round of protein synthesis. Number of members: 27
rve		Integrase core domain	Accession number: PF00665 Definition: Integrase core domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_10 (release 2.1) Gathering cutoffs: 9.3 9.3 Trusted cutoffs: 9.30 9.30 Noise cutoffs: 9.20 9.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95099322 Reference Title: Crystal structure of the catalytic domain of HIV-1 Reference Title: integrase: similarity to other polynucleotidyl transferases Reference Title: [see comments] Reference Author: Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Reference Author: Davies DR; Reference Location: Science 1994;266:1981-1986 Database Reference: SCOP; 2itg; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001584; Database Reference PDB; 1cxu A; 56; 198; Database Reference PDB; 1vsh ; 54; 199; Database Reference PDB; 1vsi ; 54; 199; Database Reference PDB; 1vsj ; 54; 199; Database Reference PDB; 1cxq A; 53; 198; Database Reference PDB; 1a5v ; 54; 199; Database Reference PDB; 1a5w ; 54; 199; Database Reference PDB; 1a5x ; 54; 199; Database Reference PDB; 1asv ; 54; 199;

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1vsm A; 54; 199;
			Database Reference PDB; 1czb A; 53; 198;
			Database Reference PDB; 1asw ; 53; 201;
			Database Reference PDB; 1cz9 A; 59; 197;
			Database Reference PDB; 1vsk ; 54; 199;
			Database Reference PDB; 1vsl A; 54; 199;
			Database Reference PDB; 1asu ; 53; 207;
			Database Reference PDB; 1c0m A; 53; 213;
			Database Reference PDB; 1vsd ; 54; 88;
			Database Reference PDB; 1vse ; 54; 88;
			Database Reference PDB; 1c1a B; 55; 213;
			Database Reference PDB; 1c0m B; 54; 213;
			Database Reference PDB; 1c0m D; 54; 213;
			Database Reference PDB; 1c1a A; 53; 213;
			Database Reference PDB; 1c0m C; 53; 213;
			Database Reference PDB; 1bhl ; 57; 201;
			Database Reference PDB; 1bi4 B; 57; 201;
			Database Reference PDB; 1bl3 B; 57; 201;
			Database Reference PDB; 1b9f A; 56; 201;
			Database Reference PDB; 1bis B; 56; 201;
			Database Reference PDB; 1qs4 B; 56; 201;
			Database Reference PDB; 1qs4 C; 56; 201;
			Database Reference PDB; 1biz A; 54; 201;
			Database Reference PDB; 1itg ; 55; 201;
			Database Reference PDB; 1bi4 C; 53; 201;
			Database Reference PDB; 1bl3 C; 53; 201;
			Database Reference PDB; 2itg ; 53; 201;
			Database Reference PDB; 1b9d A; 57; 189;
			Database Reference PDB; 1bi4 A; 57; 201;
			Database Reference PDB; 1bl3 A; 57; 201;
			Database Reference PDB; 1bis A; 56; 201;
			Database Reference PDB; 1biu A; 56; 201;
			Database Reference PDB; 1biu B; 56; 201;
			Database Reference PDB; 1biu C; 56; 201;
			Database Reference PDB; 1qs4 A; 56; 201;
			Database Reference PDB; 1b92 A; 56; 201;
			Database Reference PDB; 1biz B; 58; 201;
			Database Reference PDB; 1b9d A; 382; 390;
			Database Reference PDB; 1wjb A; 53; 55;
			Database Reference PDB; 1wjb B; 53; 55;
			Database Reference PDB; 1wjd A; 53; 55;
			Database Reference PDB; 1wjd B; 53; 55;
			Database Reference PDB; 1wjf A; 53; 55;
			Database Reference PDB; 1wjf B; 53; 55,
			Database reference: PFAMB; PB000048;
			Database reference: PFAMB; PB007709;
			Database reference: PFAMB; PB013923;
			Database reference: PFAMB; PB013938;
			Database reference: PFAMB; PB018509;
			Database reference: PFAMB; PB020302;
			Database reference: PFAMB; PB025327;
			Database reference: PFAMB; PB028352;
			Database reference: PFAMB; PB032740;
			Database reference: PFAMB; PB040612;
			Database reference: PFAMB; PB040636;
			Database reference: PFAMB; PB040684;
			Database reference: PFAMB; PB040695;
			Database reference: PFAMB; PB040730;
			Database reference: PFAMB; PB040824;
			Database reference: PFAMB; PB041112;
			Database reference: PFAMB; PB041143;
			Database reference: PFAMB; PB041275;
			Database reference: PFAMB; PB041356;
			Database reference: PFAMB; PB041375;
			Database reference: PFAMB; PB041456;
			Database reference: PFAMB; PB041459;
			Database reference: PFAMB; PB041522;
			Database reference: PFAMB; PB041665;
			Database reference: PFAMB; PB041761;
			Database reference: PFAMB; PB041816;
			Database reference: PFAMB; PB041885;
			Comment: Integrase mediates integration of a DNA copy of the viral
			Comment: genome into the host chromosome.

Pfam	Prosite	Full Name	Description
			Integrase is composed of Comment: three domains. The amino-terminal domain is a zinc binding Comment: domain Integrase_Zn. This domain is the central catalytic Comment: domain. The carboxyl terminal domain that is a non-specific Comment: DNA binding domain integrase. Comment: The catalytic domain acts as an endonuclease when two Comment: nucleotides are removed from the 3' ends of the blunt-ended Comment: viral DNA made by reverse transcription. This domain also Comment: catalyses the DNA strand transfer reaction of the 3' ends Comment: of the viral DNA to the 5' ends of the integration site [1]. Number of members: 1147
S4		S4 domain	Accession number: PF01479 Definition: S4 domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Medline:99193178 Gathering cutoffs: 17 17 Trusted cutoffs: 17.20 17.20 Noise cutoffs: 16.70 16.70 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99193178 Reference Title: Novel predicted RNA-binding domains associated with the Reference Title: translation machinery. Reference Author: Aravind L, Koonin EV; Reference Location: J Mol Evol 1999;48:291-302. Reference Number: [2] Reference Medline: 98372721 Reference Title: The crystal structure of ribosomal protein S4 reveals a Reference Title: two-domain molecule with an extensive RNA-binding surface: Reference Title: one domain shows structural homology to the ETS DNA-binding Reference Title: motif. Reference Author: Davies C, Gerstner RB, Draper DE, Ramakrishnan V, White SW; Reference Location: EMBO J 1998;17:4545-4558. Database Reference: SCOP; 1c06; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002942; Database Reference PDB; 1c05 A; 51; 98; Database Reference PDB; 1c06 A; 51; 98; Database Reference PDB; 1dm9 A; 9; 55; Database Reference PDB; 1dm9 B; 9; 55; Database reference: PFAMB; PB001751; Database reference. PFAMB; PB041147; Database reference: PFAMB; PB041148; Comment: The S4 domain is a small domain consisting of 60-65 amino acid residues Comment: that was detected in the bacterial ribosomal protein S4, eukaryotic Comment: ribosomal S9, two families of pseudouridine synthases, a novel family Comment: of predicted RNA methylases, a yeast protein containing a pseudouridine Comment: synthetase and a deaminase domain, bacterial tyrosyl-tRNA synthetases, Comment: and a number of uncharacterized, small proteins that may be involved in Comment: translation regulation [1]. The S4 domain probably mediates binding to Comment: RNA.

Pfam	Prosite	Full Name	Description
			Number of members: 256
SAA_proteins	PDOC00762	Serum amyloid A proteins signature	The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL) [1,2]. The synthesis of certain members of the family is greatly increased (as much as a 1000 fold) in inflammation; thus making SAA a major acute phase reactant. While the major physiological function of SAA is unclear, prolonged elevation of plasma SAA levels, as in chronic inflammation, however, results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. SAA are proteins of about 110 amino acid residues. As a signature pattern, we selected the most highly conserved region, which is located in the central part of the sequence. Description of pattern(s) and/or profile(s) Consensus pattern A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-G-x-W-A Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update June 1994 / First entry. References [1] Malle E., Steinmetz A., Raynes J.G. Atherosclerosis 102:131-146(1993). [2] Uhlar C.M., Burgess C.J., Sharp P.M., Whitehead A.S. Genomics 19:228-235(1994).
SAM		SAM domain (Sterile alpha motif)	Accession number: PF00536 Definition: SAM domain (Sterile alpha motif) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1],[2] Gathering cutoffs: 11 0 Trusted cutoffs: 11.00 3.70 Noise cutoffs: 10.90 10.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96100659 Reference Title: SAM: A novel motif in yeast sterile alpha and Drosophila Reference Title: polyhomeotic proteins Reference Author: Ponting CP; Reference Location: Prot Sci 1995;4:1928-1930. Reference Number: [2] Reference Medline: 97160498 Reference Title: SAM as a protein interaction domain involved in Reference Title: developmental regulation. Reference Author: Shultz J, Ponting CP, Hofmann K, Bork P; Reference Location: Prot Sci 1997;6:249-253. Reference Number: [3] Reference Medline: 99101382 Reference Title: The crystal structure of an Eph receptor SAM domain reveals Reference Title: a mechanism for modular dimerization.

Pfam	Prosite	Full Name	Description
			Reference Author: Stapleton D, Balan I, Pawson T, Sicheri F; Reference Location: Nat Struct Biol 1999;6:44-49. Database reference: SMART; SAM; Database Reference: SCOP; 1b0x; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001660; Database Reference PDB; 1b0x A; 910; 973; Database Reference PDB; 1sgg ; 7; 70; Database Reference PDB; 1b4f A; 7; 71; Database Reference PDB; 1b4f C; 7; 71; Database Reference PDB; 1b4f E; 7; 71; Database Reference PDB; 1b4f D; 7; 71; Database Reference PDB; 1b4f H; 7; 71; Database Reference PDB; 1b4f F; 7; 71; Database Reference PDB; 1b4f G; 7; 71; Database Reference PDB; 1b4f B; 7; 71; Database reference: PFAMB; PB008631; Database reference: PFAMB; PB040678; Database reference: PFAMB; PB041111; Database reference: PFAMB; PB041385; Comment: It has been suggested that SAM is an evolutionarily conserved protein Comment: binding domain that is involved in the regulation of numerous Comment: developmental processes in diverse eukaryotes. Comment: The SAM domain can potentially function as a protein interaction Comment: module through its ability to homo- and heterooligomerise with Comment: other SAM domains. Number of members: 110
SAM_decarbox		Adenosylmethionine decarboxylase	Accession number: PF01536 Definition: Adenosylmethionine decarboxylase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_600 (release 4.0) Gathering cutoffs: 11 11 Trusted cutoffs: 17.90 17.90 Noise cutoffs: 5.70 5.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98098079 Reference Title: Cloning, mapping and mutational analysis of the Reference Title: S-adenosylmethionine decarboxylase gene in Drosophila Reference Title: melanogaster. Reference Author: Larsson J, Rasmuson-Lestander A; Reference Location: Mol Gen Genet 1997;256:652-660. Database Reference: SCOP; 1jen; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001985: Database Reference PDB; 1jen C; 69; 328; Database Reference PDB; 1jen A; 69; 329; Database Reference PDB; 1jen B; 4; 67; Database Reference PDB; 1jen D; 5; 66; Comment: This is a family of S-adenosylmethionine decarboxylase (SAMDC) proenzymes. Comment: In the biosynthesis of polyamines SAMDC produces decarboxylated Comment: S-adenosylmethionine, which serves as the aminopropyl moiety necessary Comment: for spermidine and spermine biosynthesis from putrescine [1]. The Pfam Comment: alignment contains both the alpha and beta chains that are cleaved to Comment: form the active enzyme. Number of members: 34
SBF		Sodium Bile acid symporter family	Accession number: PF01758 Definition: Sodium Bile acid symporter family

Pfam	Prosite	Full Name	Description
			Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_697 (release 4.2) Gathering cutoffs: -19 -19 Trusted cutoffs: -12.50 -12.50 Noise cutoffs: -26.40 -26.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97377989 Reference Title: Isolation of three contiguous genes, ACR1, ACR2 and ACR3, Reference Title: involved in resistance to arsenic compounds in the yeast Reference Title: Saccharomyces cerevisiae. Reference Author: Bobrowicz P, Wysocki R, Owsianik G, Goffeau A, Ulaszewski Reference Author: S; Reference Location: Yeast 1997;13:819-828. Reference Number: [2] Reference Medline: 92073340 Reference Title: Functional expression cloning and characterization of the Reference Title: hepatocyte Na+/bile acid cotransport system. Reference Author: Hagenbuch B, Stieger B, Foguet M, Lubbert H, Meier PJ; Reference Location: Proc Natl Acad Sci U S A 1991;88:10629-10633. Database Reference INTERPRO; IPR002657; Database reference: PFAMB; PB041594; Comment: This family consists of Na+/bile acid co-transporters. Comment: These transmembrane proteins function in the liver Comment: in the uptake of bile acids from portal blood plasma Comment: a process mediated by the co-transport of Na+ [2]. Comment: Also in the family is ARC3 from S. cerevisiae Swiss:Q06598 Comment: this is a putative transmembrane protein involved in Comment: resistance to arsenic compounds [1]. Number of members: 29
Sec7		Sec7 domain	Accession number: PF01369 Definition: Sec7 domain Author: Bateman A Alignment method of seed: Clustalw_manual Source of seed members: Pfam-B_1629 (release 3.0) Gathering cutoffs: 25 25 Trusted cutoffs: 101.50 101.50 Noise cutoffs: 13.20 13.20 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98169075 Reference Title: Structure of the Sec7 domain of the Arf exchange factor Reference Title: ARNO. Reference Author: Cherfils J, Menetrey J, Mathieu M, Le Bras G, Robineau S, Reference Author: Beraud-Dufour S, Antonny B, Chardin P; Reference Location: Nature 1998;392:101-105. Reference Number: [2] Reference Medline: 97100951 Reference Title: A human exchange factor for ARF contains Sec7- and Reference Title: pleckstrin- homology domains. Reference Author: Chardin P, Paris S, Antonny B, Robineau S, Beraud-Dufour S, Reference Author: Jackson CL, Chabre M Reference Location: Nature 1996;384:481-484.

Pfam	Prosite	Full Name	Description
			Database Reference: SCOP; 1pbv; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000904; Database Reference PDB; 1pbv ; 58; 243; Database Reference PDB; 1bc9 ; 59; 244; Comment: The Sec7 domain is a guanine-nucleotide-exchange-factor (GEF) Comment: for the arf family [2]. Number of members: 32
Seedstore_2S		2S seed storage family	Accession number: PF01631 Definition: 2S seed storage family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1154 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 95.10 95.10 Noise cutoffs: -0.20 10.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97121264 Reference Title: 1H NMR assignment and global fold of napin BnIb, a Reference Title: representative 2S albumin seed protein. Reference Author: Rico M, Bruix M, Gonzalez C, Monsalve RI, Rodriguez R; Reference Location: Biochemistry 1996;35:15672-15682. Database Reference: SCOP; 1pnb; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR000617; Database reference: PFAMB; PB029622; Comment: Members of this family are composed of two chains (both included in Comment: the alignment), these are co-translated and later cleaved. The two Comment: chains are disulphide linked together. Number of members: 27
SH2	PDOC50001	Src homology 2 (SH2) domain profile	The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6]. The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets [7]. So far, SH2 domains have been identified in the following proteins: - Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases. - Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two copies of the SH2 domain are found in those proteins in between the catalytic 'X-' and 'Y-boxes' (see <PDOC50007>). - Mammalian phosphatidyl inositol 3-kinase regulatory p85

Pfam	Prosite	Full Name	Description
			subunit. - Some vertebrate and invertebrate protein-tyrosine phosphatases. - Mammalian Ras GTPase-activating protein (GAP). - Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK. - Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family. - Miscellanous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc. - STAT proteins (signal transducers and activators of transcription). - Chicken tensin. - Yeast transcriptional control protein SPT6. The profile developed to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions. Description of pattern(s) and/or profile(s) Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT protein tyrosine kinases JAK1 and JAK2. Expert(s) to contact by email Zvelebil M. marketa@ludwig.ucl.ac.uk Last update November 1995 / First entry. References [1] Sadowski I., Stone J.C., Pawson T. Mol. Cell. Biol. 6:4396-4408(1986). [2] Russel R.B., Breed J., Barton G.J. FEBS Lett. 304:15-20(1992). [3] Marangere L.E.M., Pawson T. J. Cell Sci. Suppl. 18:97-104(1994). [4] Pawson T., Schlessinger J. Curr. Biol. 3:434-442(1993). [5] Mayer B.J., Baltimore D. Trends Cell. Biol. 3:8-13(1993). [6] Pawson T. Nature 373:573-580(1995). [7] Kuriyan J., Cowburn D. Curr. Opin. Struct. Biol. 3:828-837(1993).
Shikimate_DH		Shikimate / quinate 5-dehydrogenase	Accession number: PF01488 Definition: Shikimate / quinate 5-dehydrogenase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_336 (release 4.0)

Pfam	Prosite	Full Name	Description
			Gathering cutoffs: -50 -50 Trusted cutoffs: -48.00 -48.00 Noise cutoffs: -82.00 -82.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96048023 Reference Title: The molecular biology of multidomain proteins. Selected Reference Title: examples. Reference Author: Hawkins AR, Lamb HK; Reference Location: Eur J Biochem 1995;232:7-18. Database Reference INTERPRO; IPR002907; Comment: This family contains both shikimate and quinate dehydrogenases. Comment: Shikimate 5-dehydrogenase catalyses the conversion of Comment: shikimate to 5-dehydroshikimate. This reaction is part of Comment: the shikimate pathway which is involved in the biosynthesis Comment: of aromatic amino acids. Comment: Quinate 5-dehydrogenase catalyses the conversion of Comment: quinate to 5-dehydroquinate. This reaction is part of Comment: the quinate pathway where quinic acid is exploited as Comment: a source of carbon in prokaryotes and microbial Comment: eukaryotes. Comment: Both the shikimate and quinate pathways share two common Comment: pathway metabolites 3-dehydroquinate and dehydroshikimate. Number of members: 58
Sigma54_factors	PDOC00593	Sigma-54 factors family signatures and profile	Sigma factors [1] are bacterial transcription initiation factors that promote the attachment of the core RNA polymerase to specific initiation sites and are then released. They alter the specificity of promoter recognition. Most bacteria express a multiplicity of sigma factors. Two of these factors, sigma- 70 (gene rpoD), generally known as the major or primary sigma factor, and sigma-54 (gene rpoN or ntrA) direct the transcription of a wide variety of genes. The other sigma factors, known as alternative sigma factors, are required for the transcription of specific subsets of genes. With regard to sequence similarity, sigma factors can be grouped into two classes: the sigma-54 and sigma-70 families. The sigma-70 family has many different sigma factors (see the relevant entry <PDOC00592>). The sigma-54 family consists exclusively of sigma-54 factor [2,3] required for the transcription of promoters that have a characteristic -24 and -12 consensus recognition element but which are devoid of the typical -10,-35 sequences recognized by the major sigma factors. The sigma-54 factor is also characterized by its interaction with ATP-dependent positive regulatory proteins that bind to upstream activating sequences. Structurally sigma-54 factors consist of three distinct regions: - A relatively well conserved N-terminal glutamine-rich region of

Pfam	Prosite	Full Name	Description
			about 50 residues that contains a potential leucine zipper motif. - A region of variable length which is not well conserved. - A well conserved C-terminal region of about 350 residues that contains a second potential leucine zipper, a potential DNA-binding 'helix-turn-helix' motif and a perfectly conserved octapeptide whose function is not known. We developed two signature patterns for this family of sigma factors. The first starts two residues before the N-terminal extremity of the helix-turn- helix region and ends two residues before its C-terminal extremity. The second is the conserved octapeptide. A profile has also been designed that covers the whole C-terminal region. Description of pattern(s) and/or profile(s) Consensus pattern P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)-[LIVMFT]-x(2)-[HS]-x- S-T-[LIVM]-S-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern R-R-T-[IV]-[ATN]-K-Y-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update July 1999 / Patterns and text revised. References [1] Helmann J.D., Chamberlin M.J. Annu. Rev. Biochem. 57:839-872(1988). [2] Thoeny B., Hennecke H. FEMS Microbiol. Rev. 5:341-358(1989). [3] Merrick M.J. Mol Microbiol. 10:903-909(1993).
SLH	PDOC00823	S-layer homology domain signature	S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain has been found in: - S-layer glycoprotein of Acetogenium kivui (3 copies). - S-layer 125 Kd protein of Bacillus sphaericus (3 copies). - S-layer protein of Bacillus anthracis (3 copies). - S-layer protein of Bacillus licheniformis (3 copies).

Pfam	Prosite	Full Name	Description
			- S-layer protein (HWP) from Bacillus brevis strain HPD31 (3 copies). - Middle cell wall protein (MWP) from Bacillus brevis strain 47 (3 copies). - S-layer protein (p100) of Thermus thermophilus (1 copy). - Outer membrane protein Omp-alpha from Thermotoga maritima (1 copy). - Cellulosome anchoring protein (gene ancA), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from Clostridium thermocellum (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation). - Amylopullulanase (gene amyB) from Thermoanaerobacter thermosulfurogenes (3 copies) - Amylopullulanase (gene aapT) from Bacillus strain XAL-601 (3 copies). - Endoglucanase from Bacillus strain KSM-635 (3 copies). - Exoglucanase (gene xynX) from Clostridium thermocellum (3 copies). - Xylanase A (gene xynA) from Thermoanaerobacter saccharolyticum (2 copies; 3 copies if a frameshift is taken into account). - Protein involved in butirosin production (ButB) from Bacillus circulans (2 incomplete copies; 3 copies if three frameshifts are taken into account). - Two hypothetical proteins from Synechocystis strain PCC 6803 (1 copy each). - A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from Bacillus circulans (fragment of 1 copy; 3 copies if two frameshifts are taken into account). SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length. Description of pattern(s) and/or profile(s) Consensus pattern [LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[FYWPDA]-x(4)-[LIV]-x(2)- [GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x- [PGAV]-x(3,10)-[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Lupas A.N. lupas@vms.biochem.mpg.de Last update November 1997 / Pattern and text revised. References [1] Beveridge T.J. Curr Opin. Struct. Biol. 4:204-212(1994). [2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,

Pfam	Prosite	Full Name	Description
			Baumeister W. J. Bacteriol. 176:1224-1233(1994). [3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).
Smr		Smr domain	Accession number: PF01713 Definition: Smr domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 0 0 Trusted cutoffs: 1.40 1.40 Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 10431172 Reference Title: Smr: a bacterial and eukaryotic homologue of the C-terminal Reference Title: region of the MutS2 family. Reference Author: Moreira D, Philippe H; Reference Location: Trends Biochem Sci 1999;24:298-300. Database Reference INTERPRO; IPR002625; Comment: This family includes the Smr (Small MutS Related) proteins, Comment: and the C-terminal region of the MutS2 protein. It has been Comment: suggested that this domain interacts with the MutS1 Comment: Swiss:P23909 protein in the case of Smr proteins and with Comment: the N-terminal MutS related region of MutS2 Swiss:P94545 [1]. Number of members: 14
SRF-TF	PDOC00302	MADS-box domain signature and profile	A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below: - Serum response factor (SRF) [1], a mammalian transcription factor that binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the transcription initiation site of genes such as c-fos. - Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. - Drosophila myocyte-specific enhancer factor 2 (MEF2). - Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes. - Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80). - Yeast transcription factor RLM1. - Yeast transcription factor SMP1. - Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower. - Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and

Pfam	Prosite	Full Name	Description
			Pistillata (PI) which act locally to specify the identity of the floral meristem and to determine sepal and petal development [4]. - Antirrhinum majus and tobacco homeotic protein deficiens (DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO cause the transformation of petals into sepals and of stamina into carpels. - Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6]. - Antirrhinum majus morphogenetic protein DEF H33 (squamosa). In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. We have derived a pattern that spans the complete length of the domain. The profile also spans the length of the MADS-box. Description of pattern(s) and/or profile(s) Consensus pattern R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)- x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x- [LIVM](3)-x(6)-[LIVMF]-x(2)-[FY] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so. Last update July 1999 / Pattern and text revised. References [1] Norman C., Runswick M., Pollock R., Treisman R. Cell 55:989-1003(1988). [2] Passmore S., Maine G.T., Elble R., Christ C., Tye B.-K. J. Mol. Biol. 204:593-606(1988). [3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. Nature 346:35-39(1990). [4] Goto K., Meyerowitz E.M. Genes Dev. 8:1548-1560(1994). [5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W.-E., Saedler H., Sommer H., Schwartz-Sommer Z. EMBO J. 11:4693-4704(1992). [6] Ma H., Yanofsky M.F., Meyerowitz E.M. Genes Dev. 5:484-495(1991). [E1] http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0014
SRP19		SRP19 protein	Accession number: PF01922 Definition: SRP19 protein Author: Enright A, Ouzounis C, Bateman A

Pfam	Prosite	Full Name	Description
			Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 31.20 31.20 Noise cutoffs: -28.50 -28.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 89041541 Reference Title: Isolation and characterization of a cDNA clone encoding the Reference Title: 19 kDa protein of signal recognition particle (SRP): Reference Title: expression and binding to 7SL RNA. Reference Author: Lingelbach K, Zwieb C, Webb JR, Marshallsay C, Hoben PJ, Reference Author: Walter P, Dobberstein B; Reference Location: Nucleic Acids Res 1988;16:9431-9442. Reference Number: [2] Reference Medline: 92220168 Reference Title: SEC65 gene product is a subunit of the yeast signal Reference Title: recognition particle required for its integrity. Reference Author: Hann BC, Stirling CJ, Walter P; Reference Location: Nature 1992;356:532-533. Reference Number: [3] Reference Medline: 92220169 Reference Title: The S. cerevisiae SEC65 gene encodes a component of yeast Reference Title: signal recognition particle with homology to human SRP19. Reference Author: Stirling CJ, Hewitt EW; Reference Location: Nature 1992;356:534-537. Database Reference INTERPRO; IPR002778; Comment: The signal recognition particle (SRP) binds to the signal peptide of Comment: proteins as they are being translated. The binding of the SRP halts Comment: translation and the complex is then transported to the endoplasmic Comment: reticulum's cytoplasmic surface. The SRP then aids translocation of Comment: the protein through the ER membrane. The SRP is a ribonucleoprotein Comment: that is composed of a small RNA and several proteins. One of these Comment: proteins is the SRP19 protein [1] (Sec65 in yeast [2,3]). Number of members: 13
SSB	PDOC00602	Single-strand binding protein family signatures	The Escherichia coli single-strand binding protein [1] (gene ssb), also known as the helix-destabilizing protein, is a protein of 177 amino acids. It binds tightly, as a homotetramer, to single-stranded DNA (ss-DNA) and plays an important role in DNA replication, recombination and repair. Closely related variants of SSB are encoded in the genome of a variety of large self-transmissible plasmids. SSB has also been characterized in bacteria such as Proteus mirabilis or Serratia marcescens. Eukaryotic mitochondrial proteins that bind ss-DNA and are probably involved in mitochondrial DNA replication are structurally and evolutionary related to prokaryotic SSB. Proteins currently known to belong to this subfamily are listed below [2]. - Mammalian protein Mt-SSB (P16). - Xenopus Mt-SSBs and Mt-SSBr.

Pfam	Prosite	Full Name	Description
			- Drosophila MtSSB. - Yeast protein RIM1. We have developed two signature patterns for these proteins. The first is a conserved region in the N-terminal section of the SSB's. The second is a centrally located region which, in Escherichia coli SSB, is known to be involved in the binding of DNA. Description of pattern(s) and/or profile(s) Consensus pattern [LIVMF]-[NST]-[KRHST]-[LIVM]-x-[LIVMF](2)-G-[NHRK]- [LIVMA]-[GST]-x-[DENT] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR] Sequences known to belong to this class detected by the pattern A majority. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Patterns and text revised. References [1] Meyer R.R., Laine P.S. Microbiol. Rev. 54:342-380(1990). [2] Stroumbakis N.D., Li Z., Tolias P.P. Gene 143:171-177(1994).
START		START domain	Accession number: PF01852 Definition: START domain Author: SMART Alignment method of seed: Manual Source of seed members: Alignment kindly provided by SMART Gathering cutoffs: 25 25 Trusted cutoffs: 106.20 106.20 Noise cutoffs: -20.90 -20.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99257451 Reference Title: START: a lipid-binding domain in StAR, HD-ZIP and Reference Title: signalling proteins. Reference Author: Ponting CP, Aravind L; Reference Location: Trends Biochem Sci 1999;24:130-132. Database reference: SMART; START; Database Reference INTERPRO; IPR002913; Number of members: 41
Sterol_desat		Sterol desaturase	Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91323727 Reference Title: Cloning, disruption and sequence of the gene encoding yeast Reference Title: C-5 sterol desaturase. Reference Author: Arthington BA, Bennett LG, Skatrud PL,

Pfam	Prosite	Full Name	Description
			Guynn CJ, Barbuch Reference Author: RJ, Ulbright CE, Bard M; Reference Location: Gene 1991;102:39-44. Reference Number: [2] Reference Medline: 96133902 Reference Title: Cloning and characterization of ERG25, the Saccharomyces Reference Title: cerevisiae gene encoding C-4 sterol methyl oxidase. Reference Author: Bard M, Bruner DA, Pierson CA, Lees ND, Biermann B, Frye L, Reference Author: Koegel C, Barbuch R; Reference Location: Proc Natl Acad Sci U S A 1996;93:186-190. Reference Number: [3] Reference Medline: 96351930 Reference Title: Molecular characterization of the CER1 gene of arabidopsis Reference Title: involved in epicuticular wax biosynthesis and pollen Reference Title: fertility. Reference Author: Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A; Reference Location: Plant Cell 1995;7:2115-2127. Database Reference INTERPRO; IPR001541; Database reference: PFAMB; PB041851; Comment: This family includes C-5 sterol desaturase and C-4 sterol methyl Comment: oxidase. Members of this family are involved in cholesterol biosynthesis Comment: and biosynthesis a plant cuticular wax. These enzymes contain many Comment: conserved histidine residues. Members of this family are integral Comment: mebrane proteins. Number of members: 34
Sulfatase	PDOC00117	Sulfatases signatures	Sulfatases (EC 3.1.6.-) are enzymes that hydrolyze various sulfate esters. The sequence of different types of sulfatases are available. These enzymes are: - Arylsulfatase A (EC 3.1.6.8) (ASA), a lysosomal enzyme which hydrolyzes cerebroside sulfate. - Arylsulfatase B (EC 3.1.6.12) (ASB), a lysosomal enzyme which hydrolyzes the sulfate ester group from N-acetylgalactosamine 4-sulfate residues of dermatan sulfate. - Arylsulfatase C (ASD). - Arylsulfatase E (ASE). - Steryl-sulfatase (EC 3.1.6.2) (STS) (arylsulfatase C), a membrane bound microsomal enzyme which hydrolyzes 3-beta-hydroxy steroid sulfates. - Iduronate 2-sulfatase precursor (EC 3.1.6.13) (IDS), a lysosomal enzyme that hydrolyzes the 2-sulfate groups from non-reducing-terminal iduronic acid residues in dermatan sulfate and heparan sulfate. - N-acetylgalactosamine-6-sulfatase (EC 3.1.6 4), an enzyme that hydrolyzes the 6-sulfate groups of the N-acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and the D-galactose 6-sulfate units of keratan sulfate. - Choline sulfatase (EC 3.1.6.6) (gene betC), a bacterial enzyme that converts choline-O-sulfate to choline. - Glucosamine-6-sulfatase (EC 3.1.6.14) (G6S), a lysosomal enzyme that hydrolyzes the N-acetyl-D-glucosamine 6-sulfate units of heparan sulfate

Pfam	Prosite	Full Name	Description
			and keratan sulfate. - N-sulphoglucosamine sulphohydrolase (EC 3.10.1.1) (sulphamidase), the lysosomal enzyme that catalyzes the hydrolysis of N-sulfo-d-glucosamine into glucosamine and sulfate. - Sea urchin embryo arylsulfatase (EC 3.1.6.1). - Green alga arylsulfatase (EC 3.1.6.1), an enzyme which plays an important role in the mineralization of sulfates. - Arylsulfatase (EC 3.1.6.1) from Escherichia coli (gene aslA), Klebsiella aerogenes (gene atsA) and Pseudomonas aeruginosa (gene atsA). - Escherichia coli hypothetical protein yidJ. It has been shown that all these sulfatases are structurally related [1,2,3]. As signature patterns for that family of enzymes we have selected the two best conserved regions. Both regions are located in the N-terminal section of these enzymes. The first region contains a conserved arginine which could be implicated in the catalytic mechanism; it is located four residues after a position that, in eukaryotic sulfatases, is a conserved cysteine which has been shown [4] to be modified to 2-amino-3-oxopropionic acid. In prokaryotes, this cysteine is replaced by a serine. Description of pattern(s) and/or profile(s) Consensus pattern [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)- [TAR]-G [R is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern G-[YV]-x-[ST]-x(2)-[IVAS]-G-K-x(0,1)-[FYWMK]-[HL] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Patterns and text revised. References [1] Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlsdorf M., Vingron M., Meyer H.E., Pohlmann R., von Figura K. J. Biol. Chem. 265:3374-3381(1990). [2] Wilson P.J., Morris C.P., Anson D.S., Occhiodoro T., Bielicki J., Clements P.R., Hopwood J.J. Proc. Natl. Acad. Sci. U.S.A. 87:8531-8535(1990). [3] de Hostos E.L., Schilling J., Grossman A.R. Mol. Gen. Genet. 218:229-239(1989). [4] Selmer T., Hallmann A., Schmidt B., Sumper M., von Figura K. Eur. J. Biochem. 238:341-345(1996).
Sulfate_transp	PDOC00870	Sulfate transporters signature	A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

Pfam	Prosite	Full Name	Description
			- Neurospora crassa sulfate permease II (gene cys-14). - Yeast sulfate permeases (genes SUL1 and SUL2). - Rat sulfate anion transporter 1 (SAT-1). - Mammalian DTDST, a probable sulfate transporter which, in Human, is involved in the genetic disease, diastrophic dysplasia (DTD). - Sulfate transporters 1, 2 and 3 from the legume Stylosanthes hamata. - Human pendrin (gene PDS), which is involved in a number of hearing loss genetic diseases. - Human protein DRA (Down-Regulated in Adenoma). - Soybean early nodulin 70. - Escherichia coli hypothetical protein ychM. - Caenorhabditis elegans hypothetical protein F41D9.5. As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]- x(3)-[GA]-[GST]-S-[KR] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Pattern and text revised. References [1] Sandal N.N., Marcker K.A. Trends Biochem. Sci. 19:19-19(1994). [2] Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T. Mol. Gen. Genet. 247:709-715(1995).
Synuclein		Synuclein	Accession number: PF01387 Definition: Synuclein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 25 25 Trusted cutoffs: 197.80 197.80 Noise cutoffs: -33.80 -33.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98424410 Reference Title: The synuclein family. Reference Author: Lavedan C; Reference Location: Genome Res 1998;8:871-880. Database Reference INTERPRO; IPR001058; Comment: There are three types of synucleins in humans, these Comment: are called alpha, beta and gamma. Alpha synuclein has Comment: been found mutated in families with autosomal dominant Comment: Parkinson's disease. A peptide of alpha synuclein has Comment: also been found in amyloid plaques in Alzheimer's Comment: patients. Number of members: 12

Pfam	Prosite	Full Name	Description
TEA	PDOC00479	TEA domain signature	The TEA domain [1,E1] is a DNA-binding region of about 66 to 68 amino acids which has been found in the N-terminal section of the following nuclear regulatory proteins: - Mammalian enhancer factor TEF-1. TEF-1 can bind to two distinct sequences in the SV40 enhancer and is a transcriptional activator. - Mammalian TEF-3, TEF-4 and TEF-5 [2], putative transcriptional activators highly similar to TEF-1. - Drosophila scalloped protein (gene sd), a probable transcription factor that functions in the regulation of cell-specific gene expression during Drosophila development, particularly in the differentiation of the nervous system [3]. - Emericella nidulans regulatory protein abaA. AbaA is involved in the regulation of conidiation (asexual spore); its expression leads to the cessation of vegetative growth. - Yeast trans-acting factor TEC1. TEC1 is involved in the activation of the Ty1 retrotransposon. - Caenorhabditis elegans hypothetical protein F28B12.2. As a signature pattern, we have used positions 39 to 67 of the TEA domain. Description of pattern(s) and/or profile(s) Consensus pattern G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-[RK](2)-Q-[LIVM]- S-S-H-[LIVM]-Q-V Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Pattern and text revised. References [1] Buerglin T.R. Cell 66:11-12(1991). [2] Jacquemin P., Hwang J.-J., Martial J.A., Dolle P., Davidson I. J. Biol. Chem. 271:21775-21785(1996). [3] Campbell S.D., Inamdar M., Rodrigues V., Raghavan V., Palazzolo M., Chovnick A. Genes Dev. 6:367-379(1992). [E1] http://transfac.gbf-braunschweig.de/cgi-bin/qt/getEntry.pl?C0024
TGT		Queuine tRNA-ribosyltransferase	Accession number: PF01702 Definition: Queuine tRNA-ribosyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1643 (release 4.1) Gathering cutoffs: -132 -132 Trusted cutoffs: -110.00 -110.00 Noise cutoffs: -155.40 -155.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96256303 Reference Title: Crystal structure of tRNA-guanine transglycosylase: RNA

Pfam	Prosite	Full Name	Description
			Reference Title: modification by base exchange. Reference Author: Romier C, Reuter K, Suck D, Ficner R; Reference Location: EMBO J 1996;15:2850-2857. Reference Number: [2] Reference Medline: 93287116 Reference Title: tRNA-guanine transglycosylase from Escherichia coli. Reference Title: Overexpression, purification and quaternary structure. Reference Author: Garcia GA, Koch KA, Chong S; Reference Location: J Mol Biol 1993;231:489-497. Database Reference: SCOP; 1pud; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002616; Database Reference PDB; 1efz A; 138; 379; Database Reference PDB; 1enu A; 138; 379; Database Reference PDB; 1pud ; 138; 379; Database Reference PDB; 1wkd ; 138; 379; Database Reference PDB; 1wke ; 138; 379; Database Reference PDB; 1wkf ; 138; 379; Database reference: PFAMB; PB037884; Comment: This is a family of queuine tRNA-ribosyltransferases Comment: EC:2.4.2.29, also known as tRNA-guanine transglycosylase Comment: and guanine insertion enzyme. Comment: Queuine tRNA-ribosyltransferase modifies tRNAs for asparagine, Comment: aspartic acid, histidine and tyrosine with queuine. Comment: It catalyses the exchange of guanine-34 at the wobble position with Comment: 7-aminomethyl-7-deazaguanine, and the addition of a cyclopentenediol Comment: moiety to 7-aminomethyl-7-deazaguanine-34 tRNA; giving a hypermodified Comment: base queuine in the wobble position [1,2]. Comment: The aligned region contains a zinc binding motif C-x-C-x2-C-x29-H, Comment: and important tRNA and 7-aminomethyl-7deazaguanine binding residues [1]. Number of members: 24
Thi4		Thi4 family	Accession number: PF01946 Definition: Thi4 family Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 526.80 526.80 Noise cutoffs: -105.00 -105.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95050223 Reference Title: Cloning, nucleotide sequence, and regulation of Reference Title: Schizosaccharomyces pombe thi4, a thiamine biosynthetic Reference Title: gene. Reference Author: Zurlinden A, Schweingruber ME; Reference Location: J Bacteriol 1994;176:6631-6635. Database Reference INTERPRO; IPR002922; Comment: This family includes Swiss:P32318 a putative thiamine biosynthetic Comment: enzyme. Number of members: 14
ThiC		ThiC family	Accession number: PF01964 Definition: ThiC family Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25

Pfam	Prosite	Full Name	Description
			Trusted cutoffs: 1047.20 1047.20 Noise cutoffs: -338.20 -338.20 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93163063 Reference Title: Structural genes for thiamine biosynthetic enzymes Reference Title: (thiCEFGH) in Escherichia coli K-12. Reference Author: Vander Horn PB, Backstrom AD, Stewart V, Begley TP; Reference Location: J Bacteriol 1993;175:982-992. Reference Number: [2] Reference Medline: 99311269 Reference Title: Thiamin biosynthesis in prokaryotes. Reference Author: Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP, Reference Author: Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ, Xi Reference Author: J; Reference Location: Arch Microbiol 1999;171:293-300. Reference Number: [3] Reference Medline: 97284509 Reference Title: Characterization of the Bacillus subtilis thiC operon Reference Title: involved in thiamine biosynthesis. Reference Author: Zhang Y, Taylor SV, Chiu HJ, Begley TP; Reference Location: J Bacteriol 1997;179:3030-3035. Database Reference INTERPRO; IPR002817; Comment: ThiC is found within the thiamine biosynthesis operon. ThiC is Comment: involved in pyrimidine biosynthesis [2]. Comment: ThiC catalyzes the substitution of the pyrophosphate of Comment: 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by Comment: 4-methyl-5-(beta-hydroxyethyl)thiazole phosphate to yield thiamine Comment: phosphate [3]. Number of members: 12
ThiJ		ThiJ/PfpI family	Accession number: PF01965 Definition: ThiJ/PfpI family Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -40.2 -40.2 Trusted cutoffs: -40.20 -40.20 Noise cutoffs: -47.00 -47.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97039868 Reference Title: The thiJ locus and its relation to phosphorylation of Reference Title: hydroxymethylpyrimidine in Escherichia coli. Reference Author: Mizote T, Tsuda M, Nakazawa T, Nakayama H; Reference Location: Microbiology 1996;142:2969-2974. Reference Number: [2] Reference Medline: 96196168 Reference Title: Sequence, expression in Escherichia coli, and analysis of Reference Title: the gene encoding a novel intracellular protease (PfpI) Reference Title: from the hyperthermophilic archaeon Pyrococcus furiosus. Reference Author: Halio SB, Blumentals II, Short SA, Merrill BM, Kelly RM; Reference Location: J Bacteriol 1996;178:2605-2612. Database Reference INTERPRO; IPR002818; Database reference: PFAMB; PB002774; Database reference: PFAMB; PB007213;

Pfam	Prosite	Full Name	Description
			Database reference: PFAMB; PB041784; Comment: This family includes ThiJ a thiamine biosynthesis Comment: enzyme [1] that catalyses the phosphorylation of Comment: hydroxymethylpyrimidine (HMP) to HMP monophosphate EC:2.7.1.49. Comment: The family also includes a the protease PfpI Swiss:Q51732 [2]. Number of members: 34
Thr_dehydrat_C		C-terminal domain of Threonine dehydratase	Accession number: PF00585 Definition: C-terminal domain of Threonine dehydratase Previous Pfam IDs: Thr_dehydratase_C; Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Bateman A Gathering cutoffs: 25 25 Trusted cutoffs: 99.90 51.30 Noise cutoffs: -1.10 -1.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98230745 Reference Title: Structure and control of pyridoxal phosphate dependent Reference Title: allosteric threonine deaminase. Reference Author: Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE, Reference Author: Chinchilla D, Eisenstein E; Reference Location: Structure 1998;6:465-475. Database Reference: SCOP; 1tdj; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001721; Database Reference PDB; 1tdj ; 424; 512; Database Reference PDB; 1tdj ; 329; 419; Comment: -!- Threonine dehydratases PALP all contain a carboxy Comment: terminal region. This region may have a regulatory role. Comment: Some members contain two copies of this region. Number of members: 30
thymidylat_synt	PDOC00086	Thymidylate synthase active site	Thymidylate synthase (EC 2.1.1.45) [1,2] catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate. Thymidylate synthase plays an essential role in DNA synthesis and is an important target for certain chemotherapeutic drugs. Thymidylate synthase is an enzyme of about 30 to 35 Kd in most species except in protozoan and plants where it exists as a bifunctional enzyme that includes a dihydrofolate reductase domain. A cysteine residue is involved in the catalytic mechanism (it covalently binds the 5,6-dihydro-dUMP intermediate). The sequence around the active site of this enzyme is conserved from phages to vertebrates. Description of pattern(s) and/or profile(s) Consensus pattern R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]- x(3)-[QMT]-[FYW]-x-[LV] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.

Pfam	Prosite	Full Name	Description
			Last update November 1997 / Pattern and text revised. References [1] Benkovic S.J. Annu. Rev. Biochem. 49:227-251(1980). [2] Ross P., O'Gara F., Condon S. Appl. Environ. Microbiol. 56:2156-2163(1990).
Top6A		Type II DNA topoisomerase	Accession number: PF01962 Definition: Type II DNA topoisomerase Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -99 -99 Trusted cutoffs: -40.40 -40.40 Noise cutoffs: -158.40 -158.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97238688 Reference Title: An atypical topoisomerase II from Archaea with implications Reference Title: for meiotic recombination [see comments] Reference Author: Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Reference Author: Forterre P; Reference Location: Nature 1997;386:414-417. Database Reference: SCOP; 1d3y; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002815; Database Reference PDB; 1d3y A; 77; 363; Database Reference PDB; 1d3y B; 77; 363; Comment: Members of this family are the A subunit from type II DNA Comment: topoisomerases. Type II DNA topoisomerases catalyse the relaxation Comment: of DNA supercoiling by causing transient double strand breaks. Comment: The family includes topoisomerase VI subunit A from archaebacteria Comment: Swiss:Q57815 EC:5.99.1.3 and SPO11 from yeast Swiss:P23179. Comment: A conserved tyrosine is thought to be involved in breaking the Comment: double stranded DNA [1]. Number of members: 9
Topoisom_bac	PDOC00333	Prokaryotic DNA topoisomerase I active site	DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type I topoisomerases act by catalyzing the transient breakage of DNA, one strand at a time, and the subsequent rejoining of the strands. When a prokaryotic type I topoisomerase breaks a DNA backbone bond, it simultaneously forms a protein- DNA link where the hydroxyl group of a tyrosine residue is joined to a 5'- phosphate on DNA, at one end of the enzyme-severed DNA strand. Prokaryotic organisms, such as Escherichia coli, have two type I topoisomerase isozymes: topoisomerase I (gene topA) and topoisomerase III (gene topB). Eukaroytes also contain homologs of prokaryotic topoisomerase III. There are a number of conserved residues in the region around the active site

Pfam	Prosite	Full Name	Description
			tyrosine; we used this region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [EQ]-x-L-Y-[DEQST]-x(3,12)-[LIV]-[ST]-Y-x-R-[ST]-[DEQS] [The second Y is the active site tyrosine] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990). [2] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995). [3] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991). [4] Roca J. Trends Biochem. Sci. 20:156-160(1995). [E1] http://ellington.pharm.arizona.edu/~bear/top/topo.html
toxin_3		long chain scorpion toxins	Accession number: PF00537 Definition: long chain scorpion toxins Author: Bateman A Alignment method of seed: Manual Source of seed members: Arne Elofsson. Gathering cutoffs: 25 25 Trusted cutoffs: 59.50 59.50 Noise cutoffs: -3.80 -3.80 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference: SCOP; 2sn3; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002061; Comment: -!- Scorpion toxins bind to sodium channels and inhibit the activation Comment: mechanisms of the channels, thereby blocking neuronal transmission. Number of members: 77
Translin		Translin family	Accession number: PF01997 Definition: Translin family Previous Pfam IDs: DUF130; Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 299.50 299.50 Noise cutoffs: -72.40 -72.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97165975 Reference Title: Isolation and characterization of a cDNA encoding a Reference Title: Translin-like protein, TRAX. Reference Author: Aoki K, Ishida R, Kasai M; Reference Location: FEBS Lett 1997;401:109-112. Database Reference INTERPRO; IPR002848; Comment: Members of this family include Translin Swiss:Q15631 that interacts Comment: with DNA and forms a ring around the DNA.

Pfam	Prosite	Full Name	Description
			This family also includes Comment: Swiss:Q99598, that was found to interact with translin with yeast Comment: two-hybrid screen [1]. Number of members: 10
Transposase_19		Transposase 19	Members of this family are capable of in vitro and/or in vivo insertion of a donor polynucleotide into a target polynucleotide. Such biological activity is useful for inserting DNA into host genome, for example, for cloning purposes to generate a desired vector in vitro.
TRANSPOSASE_IS 30	PDOC00801	Transposases, IS30 family, signature	Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2] to consist of transposases from the following elements: - Is30 from Escherichia coli. - Is1086 from Alcaligenes eutrophus. - Is1161 from Streptococcus salivarius. - Is4351 (Tn4551) from Bacteroides fragilis. These transposases are proteins of 340 to 380 amino acids. The best conserved region is located in their C-terminal section and is used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1995 / First entry. References [1] Dong Q., Sadouk A., van der Lelie D., Taghavi S., Ferhat A., Nuyten J.M., Borremans B., Mergeay M., Toussaint A. J. Bacteriol. 174:8133-8138(1992). [2] Giffard P.M., Rathsam C., Kwan E., Kwan D.W.L., Bunny K.L., Koo S.-P., Jacques N.A. J. Gen. Microbiol. 139:913-920(1993).
Transthyretin	PDOC00617	Transthyretin signatures	Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis. The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family: - Escherichia coli hypothetical protein yedX. - Bacillus subtilis hypothetical protein yunM.

Pfam	Prosite	Full Name	Description
			- Caenorhabditis elegans hypothetical protein R09H10.3. - Caenorhabditis elegans hypothetical protein ZK697.8. We selected two regions as signature patterns. The first located in the N- terminal extremity starts with a lysine known to be involved in binding T4. The second pattern is located in the C-terminal extremity. Description of pattern(s) and/or profile(s) Consensus pattern [KH]-[IV]-L-[DN]-x(3)-G-x-P-A-x(2)-[IV]-x-[IV] [The K binds thyroxine] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern Y-[TH]-[IV]-[AP]-x(2)-L-S-[PQ]-[FYW]-[GS]-[FY]-[QS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / Patterns and text revised. References [1] Schreiber G., Richardson S.J. Comp. Biochem. Physiol. 116B:137-160(1997).
TRM		N2,N2-dimethylguanosine tRNA methyltransferase	Accession number: PF02005 Definition: N2,N2-dimethylguanosine tRNA methyltransferase Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 664.60 664.60 Noise cutoffs: -259.50 -259.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98352211 Reference Title: The tRNA(guanine-26,N2-N2) methyltransferase (Trm1) from Reference Title: the hyperthermophilic archaeon Pyrococcus furiosus: Reference Title: cloning, sequencing of the gene and its expression in Reference Title: Escherichia coli. Reference Author: Constantinesco F, Benachenhou N, Motorin Y, Grosjean H; Reference Location: Nucleic Acids Res 1998;26:3753-3761. Reference Number: [2] Reference Medline: 87260951 Reference Title: Amino-terminal extension generated from an upstream AUG Reference Title: codon is not required for mitochondrial import of yeast Reference Title: N2,N2-dimethylguanosine- specific tRNA methyltransferase. Reference Author: Ellis SR, Hopper AK, Martin NC; Reference Location: Proc Natl Acad Sci U S A 1987;84:5172-5176. Database Reference INTERPRO; IPR002905; Database reference: PFAMB; PB041661; Comment: This enzyme EC:2.1.1.32 used S-AdoMet to methylate tRNA. Comment: The TRM1 gene of Saccharomyces cerevisiae is necessary for Comment: the N2,N2-dimethylguanosine modification of both mitochondrial Comment: and cytoplasmic tRNAs [1]. The enzyme is

Pfam	Prosite	Full Name	Description
			found in both Comment: eukaryotes and archaebacteria [2] Number of members: 10
tRNA_bind		Putative tRNA binding domain	Accession number: PF01588 Definition: Putative tRNA binding domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_482 (release 4.1) Gathering cutoffs: 20 20 Trusted cutoffs: 22.30 22.30 Noise cutoffs: 18.20 18.20 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97306356 Reference Title: Human tyrosyl-tRNA synthetase shares amino acid sequence Reference Title: homology with a putative cytokine. Reference Author: Kleeman TA, Wei D, Simpson KL, First EA; Reference Location: J Biol Chem 1997;272:14420-14425. Reference Number: [2] Reference Medline: 97050848 Reference Title: The yeast protein Arc1p binds to tRNA and functions as a Reference Title: cofactor for the methionyl-and glutamyl-tRNA synthetases. Reference Author: Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A, Reference Author: Mann M, Hurt EC; Reference Location: EMBO J 1996;15:5437-5448. Database Reference: SCOP; 1pys; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002547; Database Reference PDB; 1b70 B; 153; 247; Database Reference PDB; 1b7y B; 153; 247; Database Reference PDB; 1eiy B; 153; 247; Database Reference PDB; 1pys B; 153; 247; Database reference: PFAMB; PB010015; Comment. This domain is found in prokaryotic methionyl-tRNA synthetases, Comment: prokaryotic phenylalanyl tRNA synthetases the yeast GU4 nucleic-binding Comment: protein (G4p1 or p42, ARC1) [2], human tyrosyl-tRNA synthetase [1], Comment: and endothelial-monocyte activating polypeptide II. Comment: G4p1 binds specifically to tRNA form a complex with methionyl-tRNA Comment: synthetases [2]. In human tyrosyl-tRNA synthetase this domain may direct Comment: tRNA to the active site of the enzyme [2]. This domain may perform a Comment: common function in tRNA aminoacylation [1]. Number of members: 46
tRNA-synt_2d	PDOC00363	Aminoacyl-transfer RNA synthetases class-II signatures	Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure.

Pfam	Prosite	Full Name	Description
			The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. We have derived signature patterns from two of these regions. Description of pattern(s) and/or profile(s) Consensus pattern [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE] Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for alanine, glycine as well as bacterial histidine. Other sequence(s) detected in SWISS-PROT 43. Consensus pattern [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x- [LIVMSTAG]-[LIVMFY] Sequences known to belong to this class detected by the pattern the majority of class-II tRNA synthetases with the exception of those specific for serine and proline. Other sequence(s) detected in SWISS-PROT 161. Expert(s) to contact by email Cusack S. cusack@embl-grenoble.fr Last update July 1998 / Text revised. References [1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987). [2] Delarue M., Moras D. BioEssays 15:675-687(1993). [3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991). [4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991). [5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991). [6] Cusack S. Biochimie 75:1077-1081(1993). [7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347.249-255(1990). [8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).
trypsin	PDOC00124	Serine proteases, trypsin family, active sites	The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-

Pfam	Prosite	Full Name	Description
			bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases [1]. A partial list of proteases known to belong to the trypsin family is shown below. - Acrosin. - Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C. - Cathepsin G. - Chymotrypsins. - Complement components C1r, C1s, C2, and complement factors B, D and I. - Complement-activating component of RA-reactive factor. - Cytotoxic cell proteases (granzymes A to H). - Duodenase I. - Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin). - Enterokinase (EC 3.4.21.9) (enteropeptidase). - Hepatocyte growth factor activator. - Hepsin. - Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin). - Plasma kallikrein. - Mast cell proteases (MCP) 1 (chymase) to 8. - Myeloblastin (proteinase 3) (Wegener's autoantigen). - Plasminogen activators (urokinase-type, and tissue-type). - Trypsins I, II, III, and IV. - Tryptases. - Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator. - Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab. - Apolipoprotein(a). - Blood fluke cercarial protease. - Drosophila trypsin like proteases: alpha, easter, snake-locus. - Drosophila protease stubble (gene sb). - Major mite fecal allergen Der p III. All the above proteins belong to family S1 in the classification of peptidases [2,E1] and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases are listed below. - Achromobacter lyticus protease I. - Lysobacter alpha-lytic protease. - Streptogrisin A and B (Streptomyces proteases A and B). - Streptomyces griseus glutamyl endopeptidase II. - Streptomyces fradiae proteases 1 and 2. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for complement components C1r and C1s, pig plasminogen, bovine protein C, rodent urokinase, ancrod, gyroxin and two insect trypsins. Other sequence(s) detected in SWISS-PROT 14.

Pfam	Prosite	Full Name	Description
			Consensus pattern [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for 18 different proteases which have lost the first conserved glycine. Other sequence(s) detected in SWISS-PROT H.influenzae protease HAP which belongs to family S6 and 3 other proteins. Note if a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100% Last update November 1997 / Text revised. References [1] Brenner S. Nature 334:528-530(1988). [2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
TYA		TYA transposon protein	Accession number: PF01021 Definition: TYA transposon protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_90 (release 3.0) Gathering cutoffs: 15 15 Trusted cutoffs: 18.00 18.00 Noise cutoffs: 13.70 13.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97404699 Reference Title: Cryo-electron microscopy structure of yeast Ty Reference Title: retrotransposon virus-like particles. Reference Author: Palmer KJ, Tichelaar W, Myers N, Burns NR, Butcher SJ, Reference Author: Kingsman AJ, Fuller SD, Saibil HR; Reference Location: J Virol 1997;71:6863-6868. Database Reference INTERPRO; IPR001042; Comment: Ty are yeast transposons. A 5.7kb transcript codes Comment: for p3 a fusion protein of TYA and TYB. The TYA Comment: protein is analogous to the gag protein of retroviruses. Comment: TYA a is cleaved to form 46kd protein which can form Comment: mature virion like particles [1]. Number of members: 62
tyrosinase	PDOC00398	Tyrosinase signatures	Tyrosinase (EC 1.14.18.1) [1] is a copper monooxygenases that catalyzes the hydroxylation of monophenols and the oxidation of o-diphenols to o-quinols. This enzyme, found in prokaryotes as well as in eukaryotes, is involved in the formation of pigments such as melanins and other polyphenolic compounds. Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ion has been shown [2] to be bound by three conserved histidines residues. The regions around these copper-binding ligands are well conserved and also shared by some hemocyanins, which are copper-containing oxygen carriers from the hemolymph of

Pfam	Prosite	Full Name	Description
			many molluscs and arthropods [3,4]. At least two proteins related to tyrosinase are known to exist in mammals: - TRP-1 (TYRP1) [5], which is responsible for the conversion of 5,6-dihydro- xyindole-2-carboxylic acid (DHICA) to indole-5,6-quinone-2-carboxylic acid. - TRP-2 (TYRP2) [6], which is the melanogenic enzyme DOPAchrome tautomerase (EC 5.3.3.12) that catalyzes the conversion of DOPAchrome to DHICA. TRP-2 differs from tyrosinases and TRP-1 in that it binds two zinc ions instead of copper [7]. Other proteins that belong to this family are: - Plants polyphenol oxidases (PPO) (EC 1.10.3.1) which catalyze the oxidation of mono- and o-diphenols to o-diquinones [8]. - Caenorhabditis elegans hypothetical protein C02C2.1. We have derived two signature patterns for tyrosinase and related proteins. The first one contains two of the histidines that bind CuA, and is located in the N-terminal section of tyrosinase. The second pattern contains a histidine that binds CuB, that pattern is located in the central section of the enzyme. Description of pattern(s) and/or profile(s) Consensus pattern H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LVM]-x(3)-E [The two H's are copper ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D [H is a copper ligand] Sequences known to belong to this class detected by the pattern ALL the tyrosinases as well as all the hemocyanins. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Patterns and text revised. References [1] Lerch K. Prog. Clin. Biol. Res. 256:85-98(1988). [2] Jackman M.P., Hajnal A., Lerch K. Biochem. J. 274:707-713(1991). [3] Linzen B. Naturwissenschaften 76:206-211(1989). [4] Lang W.H., van Holde K.E. Proc. Natl. Acad. Sci. U.S.A. 88:244-248(1991). [5] Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borron J.C., Hearing V.J. EMBO J. 13:5818-5825(1994). [6]

Pfam	Prosite	Full Name	Description
			Jackson I.J., Chambers D.M., Tsukamoto K., Copeland N.G., Gilbert D.J., Jenkins N.A., Hearing V. EMBO J. 11:527-535(1992). [7] Solano F., Martinez-Liarte J.H., Jimenez-Cervantes C., Garcia-Borron J.C., Lozano J.A. Biochem. Biophys. Res. Commun. 204:1243-1250(1994). [8] Cary J.W , Lax A.R., Flurkey W.H. Plant Mol. Biol. 20:245-253(1992).
UbiA	PDOC00727	UbiA prenyltransferase family signature	The following prenyltransferases are evolutionary related [1,2]: - Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA). - Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene COQ2). - Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene COX10) and from bacteria (genes cyoE or ctaB). These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and we used it as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern N-x(3)-[DEH]-x(2)-[LIMF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)- G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Melzer M., Heide L. Biochim. Biophys. Acta 1212:93-102(1994). [2] Mogi T., Saiki K., Anraku Y. Mol. Microbiol. 14:391-398(1994).
Ubie_methyltran	PDOC00911	ubiE/COQ5 methyltransferase family signatures	The following methyltransferases have been shown [1] to share regions of similarities: - Escherichia coli ubiE, which is involved in both ubiquinone and menaquinone biosynthesis and which catalyzes the S-adenosylmethionine dependent methylation of 2-polyprenyl-6-methoxy-1,4-benzoquinol into 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol and of demethylmenaquinol into menaquinol. - Yeast COQ5, a ubiquinone biosynthesis methlytransferase. - Bacillus subtilis spore germination protein C2 (gene: gercB or gerC2), a probable menaquinone biosynthesis methlytransferase. - Lactococcus lactis gerC2 homolog. - Caenorhabditis elegans hypothetical protein ZK652.9. - Leishmania donovani amastigote-specific protein A41. These are hydrophilic proteins of about 30 Kd (except for ZK652.9 which is 65 Kd). They can be picked up in the database by the following patterns.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern R-V-[LIVM]-K-[PV]-[GM]-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Lee P.T., Hsu A.Y., Ha H.T., Clarke C.F. J. Bacteriol. 179:1748-1754(1997).
ubiquitin	PDOC00271	Ubiquitin domain signature and profile	Ubiquitin [1,2,3] is a protein of seventy six amino acid residues, found in all eukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis. In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable (up to twelve in a Xenopus gene). In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat. The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins. Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg- Gly-Gly) extending from the compact structure to form a 'tail', important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins. There are a number of proteins which are evolutionary related to ubiquitin: - Ubiquitin-like proteins from baculoviruses as well as in some strains of bovine viral diarrhea viruses (BVDV). These proteins are highly similar to their eukaryotic counterparts. - Mammalian protein GDX [4]. GDX is composed of two domains, a N-terminal ubiquitin-like domain of 74 residues and a C-terminal domain of 83 residues with some similarity with the thyroglobulin hormonogenic site. - Mammalian protein FAU [5]. FAU is a fusion protein which

Pfam	Prosite	Full Name	Description
			consist of a N-terminal ubiquitin-like protein of 74 residues fused to ribosomal protein S30. - Mouse protein NEDD-8 [6], a ubiquitin-like protein of 81 residues. - Human protein BAT3, a large fusion protein of 1132 residues that contains a N-terminal ubiquitin-like domain. - Caenorhabditis elegans protein ubl-1 [7]. Ubl-1 is a fusion protein which consist of a N-terminal ubiquitin-like protein of 70 residues fused to ribosomal protein S27A. - Yeast DNA repair protein RAD23 [8]. RAD23 contains a N-terminal domain that seems to be distantly, yet significantly, related to ubiquitin. - Mammalian RAD23-related proteins RAD23A and RAD23B. - Mammalian BCL-2 binding athanogene-1 (BAG-1). BAG-1 is a protein of 274 residues that contains a central ubiquitin-like domain. - Human spliceosome associated protein 114 (SAP 114 or SF3A120). - Yeast protein DSK2, a protein involved in spindle pole body duplication and which contains a N-terminal ubiquitin-like domain. - Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain a N-terminal ubiquitin domain and a C-terminal CAP-Gly domain (see <PDOC00660>). - Schizosaccharomyces pombe hypothetical protein SpAC26A3.16. This protein contains a N-terminal ubiquitin domain. - Yeast protein SMT3. - Human ubiquitin-like proteins SMT3A and SMT3B. - Human ubiquitin-like protein SMT3C (also known as PIC1; Ubl1, Sumo-1, Gmp-1 or Sentrin). This protein is involved in targeting ranGAP1 to the nuclear pore complex protein ranBP2. - SMT3-like proteins in plants and Caenorhabditis elegans. To identify ubiquitin and related proteins we have developed a pattern based on conserved positions in the central section of the sequence. A profile was also developed that spans the complete length of the ubiquitin domain. Description of pattern(s) and/or profile(s) Consensus pattern K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]- [LIVMC]-[LIVMFY]-x-G-x(4)-[DE] Sequences known to belong to this class detected by the pattern ALL, except for the RAD23 and SMT3 subfamilies, BAG-1 and SAP 114. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE. Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Last update July 1998 / Text revised.

Pfam	Prosite	Full Name	Description
			Bio/Technology 8:209-215(1990). References [1] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991). [2] Monia B.P., Ecker D.J., Croke S.T. [3] Finley D., Varshavsky A. Trends Biochem. Sci. 10:343-347(1985). [4] Filippi M., Tribioli C., Toniolo D. Genomics 7:453-457(1990). [5] Olvera J., Wool I.G. J. Biol. Chem. 268:17967-17974(1993). [6] Kumar S., Yoshida Y., Noda M. Biochem. Biophys. Res. Commun. 195:393-399(1993). [7] Jones D., Candido E.P. J. Biol. Chem. 268:19545-19551(1993). [8] Melnick L., Sherman F. J. Mol. Biol. 233:372-388(1993).
UPF0004	PDOC00984	Uncharacterized protein family UPF0004 signature	The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yliG. - Escherichia coli hypothetical protein yleA and HI0019, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yqeV. - Helicobacter pylori hypothetical protein HP0269. - Helicobacter pylori hypothetical protein HP0285. - Mycoplasma iowae hypothetical protein in 16S RNA 5'region. - Mycobacterium tuberculosis hypothetical protein Rv2733c. - Rickettsia prowazekii hypothetical protein RP416. - Rickettsia prowazekii hypothetical protein RP808. - Synechocystis strain PCC 6803 hypothetical protein slr0082. - Synechocystis strain PCC 6803 hypothetical protein sll0996. - Methanococcus jannaschii hypothetical protein MJ0865. - Methanococcus jannaschii hypothetical protein MJ0867. - Caenorhabditis elegans hypothetical protein F25B5.5. The size of these proteins range from 47 to 61 Kd. They contain six conserved cysteines, three of which are clustered in a region that can be used as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVMT]- x(4)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update December 1999 / Pattern and text revised. References [1] Bairoch A.

Pfam	Prosite	Full Name	Description
			Unpublished observations (1997).
UPF0013		Uncharacterized membrane protein family UPF0013	Accession number: PF01554 Definition: Uncharacterized membrane protein family UPF0013 Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_163 (release 4.0) Gathering cutoffs: -26 -26 Trusted cutoffs: -16.10 -16.10 Noise cutoffs: -36.70 -36.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference: URL; http://www.expasy.ch/cgi-bin/lists?upflist.txt; Database Reference INTERPRO; IPR002528; Database reference: PFAMB; PB041103; Comment: These proteins are integral membrane proteins of unknown Comment: function. Number of members: 47
UPF0019	PDOC00949	Uncharacterized protein family UPF0019 signature	The following uncharacterized proteins have been shown [1,2] to be highly similar: - Yeast protein SNZ1, which may be involved in growth arrest and cellular response to nutrient limitation. - Yeast chromosome VI hypothetical protein YFL059w. - Yeast chromosome XIV hypothetical protein YNL333w. - Fission yeast hypothetical protein SpAC29B12.04. - Hevea brasiliensis ethylene-inducible protein HEVER. - Stellaria longipes hypothetical protein H47. - Bacillus subtilis hypothetical protein yaaD. - Haemophilus influenzae hypothetical protein HI1647. - Mycobacterium leprae hypothetical protein MICL581.12c. - Mycobacterium tuberculosis hypothetical protein MtCY1A10.27. - Archaeoglobus fulgidus hypothetical protein AF0508. - Methanococcus jannaschii hypothetical protein MJ0677. - Methanococcus vannielii hypothetical protein in tRNA/5S rRNA gene cluster. - Methanobacterium thermoautotrophicum hypothetical protein Mth666. These are hydrophilic proteins of about 32 Kd. They can be picked up in the database by the following pattern. Description of pattern(s) and/or profile(s) Consensus pattern L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Pattern and text revised. References [1] Sivasubramaniam S., Vanniasingham V.M., Tan C.T., Chua N.H. Plant Mol. Biol. 29:173-178(1995). [2] Braun E.L., Fuge E.K., Padilla P.A., Werner-Washburne M. J. Bacteriol. 178:6865-6872(1996).
UPF0047	PDOC01018	Uncharacterized protein family UPF0047 signature	The following uncharacterized proteins have been shown [1] to be highly similar: - Bacillus subtilis hypothetical protein yugU.

Pfam	Prosite	Full Name	Description
			- Escherichia coli hypothetical protein yjbQ. - Mycobacterium tuberculosis hypothetical protein MtCY9C4.12. - Synechocystis strain PCC 6803 hypothetical protein sll1880. - Archaeoglobus fulgidus hypothetical protein AF2050. - Methanococcus jannaschii hypothetical protein MJ1081. - Methanobacterium thermoautotrophicum hypothetical protein MTH771. - Fission yeast hypothetical protein SpAC4A8.02c. These are small proteins of 14 to 16 Kd. They can be picked up in the database by the following pattern. This pattern is located in the C-terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / First entry. References [1] Bairoch A. Unpublished observations (1998).
UPF0052		Uncharacterised protein family UPF0052	Accession number: PF01933 Definition: Uncharacterised protein family UPF0052 Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 263.90 263.90 Noise cutoffs: -134.40 -134.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002882; Number of members: 12
UPF0057	PDOC01013	Uncharacterized protein family UPF0057 signature	The following uncharacterized proteins have been shown [1] to be evolutionary related: - Barley low-temperature induced protein blt101. - Lophorium elongatum salt-sress induced protein ESI3. - Yeast hypothetical proteins YDL123w, YDR276c, YDR525Bw and YJL151c. - Caenorhabditis elegans hypothetical proteins F47B7.1, T23F2.3, T23F2.4, T23F2.5 and ZK632.10. - Escherichia coli hypothetical protein yqaE. - Synechocystis strain PCC 6803 hypothetical protein ssr1169. These are small proteins of from 52 to 140 amino-acid resiudes that contains two transmembrane domains. As a signature pattern we selected a region that corresponds to the end of the first transmembrane helix. Description of pattern(s) and/or profile(s) Consensus pattern [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update

Pfam	Prosite	Full Name	Description
			July 1998 / First entry. References [1] Rudd K.E., Humphery-Smith I., Wasinger V.C., Bairoch A. Electrophoresis 19:536-544(1998).
UPF0066	PDOC01022	Uncharacterized protein family UPF0066 signature	The following uncharacterized proteins have been shown [1] to be evolutionary related: - Escherichia coli hypothetical protein yaeB and HI0510, the corresponding Haemophilus influenzae protein. - Agrobacterium tumefaciens Ti plasmid protein virR. - Pseudomonas aeruginosa protein rcsF. - Archaeoglobus fulgidus hypothetical protein AF0241. - Archaeoglobus fulgidus hypothetical protein AF0433. - Methanococcus jannaschii hypothetical protein MJ1583. - Methanobacterium thermoautotrophicum hypothetical protein MTH1797. These are proteins of from 120 to 240 amino-acid resiudes (with the exception of AF0433 which is 366 residues long). As a signature pattern we selected a conserved region in the central part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern G-[AV]-F-[STA]-x-R-[SA]-x(2)-R-P-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1999 / First entry. References [1] Bairoch A. Unpublished observations (1998).
UPF0076	PDOC00838	Uncharacterized protein family UPF0076 signature	The following uncharacterized proteins have been shown [1] to share regions of similarities: - Goat antigen UK114, a human homolog and the rat corresponding protein which is known as perchloric acid soluble protein (PSP1). PSP1 [2] may inhibit an initiation stage of cell-free protein synthesis. - Mouse heat-responsive protein HRSP12. - Yeast chromosome V hypothetical protein YER057c. - Yeast chromosome IX hypothetical protein YIL051c. - Caenorhabditis elegans hypothetical protein C23G10.2. - Escherichia coli hypothetical protein ycdK. - Escherichia coli hypothetical protein yhaR. - Escherichia coli hypothetical protein yjgF and HI0719, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yoaB. - Bacillus subtilis hypothetical protein yabJ. - Haemophilus influenzae hypothetical protein HI1627. - Helicobacter pylori hypothetical protein HP0944. - Lactococcus lactis aldR. - Myxococcus xanthus dfrA. - Synechocystis strain PCC 6803 hypothetical protein slr0709. - Rhizobium strain NGR234 symbiotic plasmid hypothetical protein y4sK. - Pyrococcus horikoshii hypothetical protein PH0854. These are small proteins of around 15 Kd whose sequence is highly conserved. As a signature pattern, we selected a well conserved region

Pfam	Prosite	Full Name	Description
			located in the C- terminal part of these proteins. Description of pattern(s) and/or profile(s) Consensus pattern [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)-[GSAKR]-x-[LMVA]- x(5,8)-[LIVM]-E-[MI] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4. Last update July 1999 / Pattern and text revised. References [1] Bairoch A. Unpublished observations (1995). [2] Oka T., Tsuji H., Noda C., Sakai K., Hong Y.-M., Suzuki I., Munoz S., Natori Y. J. Biol. Chem. 270:30060-30067(1995).
UPF0099		Domain of unknown function UPF0099	Accession number: PF01981 Definition: Domain of unknown function UPF0099 Previous Pfam IDs: DUF119; Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: 25 25 Trusted cutoffs: 132.80 132.80 Noise cutoffs: -35.70 -35.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference INTERPRO; IPR002833; Comment: This domain has no known function. Number of members: 10
UQ_con	PDOC00163	Ubiquitin-conjugating enzymes active site	Ubiquitin-conjugating enzymes (EC 6.3.2.19) (UBC or E2 enzymes) [1,2,3] catalyze the covalent attachment of ubiquitin to target proteins. An activated ubiquitin moiety is transferred from an ubiquitin-activating enzyme (E1) to E2 which later ligates ubiquitin directly to substrate proteins with or without the assistance of 'N-end' recognizing proteins (E3). In most species there are many forms of UBC (at least 9 in yeast) which are implicated in diverse cellular functions. A cysteine residue is required for ubiquitin-thiolester formation. There is a single conserved cysteine in UBC's and the region around that residue is conserved in the sequence of known UBC isozymes. We have used that region as a signature pattern. Description of pattern(s) and/or profile(s) Consensus pattern [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x- [LIV] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast UBC6 (DOA2). Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Jentsch S. jentsch@zmbh.uni-heidelberg.de Last update

Pfam	Prosite	Full Name	Description
			July 1998 / Text revised. References [1] Jentsch S., Seufert W., Sommer T , Reins H.-A. Trends Biochem. Sci. 15:195-198(1990). [2] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991). [3] Hershko A. Trends Biochem. Sci. 16:265-268(1991).
urease_gamma	PDOC00133	Urease signatures	Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma). Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3]. As signatures for this enzyme, we selected a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine. Description of pattern(s) and/or profile(s) Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988). [2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989). [3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).
UreD		UreD urease accessory protein	Accession number: PF01774 Definition: UreD urease accessory protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1109 (release 4.2) Gathering cutoffs: 25 25

Pfam	Prosite	Full Name	Description
			Trusted cutoffs: 186.00 186.00 Noise cutoffs: -42.60 -42.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97352660 Reference Title: Characterization of UreG, identification of a Reference Title: UreD-UreF-UreG complex, and evidence suggesting that a Reference Title: nucleotide-binding site in UreG is required for in vivo Reference Title: metallocenter assembly of Klebsiella aerogenes urease. Reference Author: Moncrief MB, Hausinger RP; Reference Location: J Bacteriol 1997;179:4081-4086. Reference Number: [2] Reference Medline: 96146510 Reference Title: Organization of Ureaplasma urealyticum urease gene cluster Reference Title: and expression in a suppressor strain of Escherichia coli. Reference Author: Neyrolles O, Ferris S, Behbahani N, Montagnier L, Blanchard Reference Author: A; Reference Location: J Bacteriol 1996;178:647-655. Reference Number: [3] Reference Medline: 94211837 Reference Title: In vitro activation of urease apoprotein and role of UreD Reference Title: as a chaperone required for nickel metallocenter assembly. Reference Author: Park IS, Carr MB, Hausinger RP; Reference Location: Proc Natl Acad Sci U S A 1994;91:3233-3237. Database Reference INTERPRO; IPR002669; Comment: UreD is a urease accessory protein. Urease urease hydrolyses Comment: urea into ammonia and carbamic acid [2]. UreD is involved in Comment: activation of the urease enzyme via the UreD-UreF-UreG-urease complex Comment: [1] and is required for urease nickel metallocenter assembly [3]. Comment: See also UreF UreF, UreG HypB_UreG. Number of members: 23
UreF		UreF	Accession number: PF01730 Definition: UreF Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_2037 (release 4.1) Gathering cutoffs: -31 -31 Trusted cutoffs: -14.30 -14.30 Noise cutoffs: -49.30 -49.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96404789 Reference Title: Purification and activation properties of UreD-UreF-urease Reference Title: apoprotein complexes. Reference Author: Moncrief MB, Hausinger RP; Reference Location: J Bacteriol 1996;178:5417-5421. Reference Number: [2] Reference Medline: 96146510 Reference Title: Organization of Ureaplasma urealyticum urease gene cluster Reference Title: and expression in a suppressor strain of Escherichia coli. Reference Author: Neyrolles O, Ferris S, Behbahani N, Montagnier L, Blanchard Reference Author: A; Reference Location: J Bacteriol 1996;178:647-655. Database Reference INTERPRO; IPR002639;

Pfam	Prosite	Full Name	Description
			Comment: This family consists of the Urease accessory protein Comment: UreF. The urease enzyme (urea amidohydrolase) Comment: hydrolyses urea into ammonia and carbamic acid [2]. Comment: UreF is proposed to modulate the activation process of Comment: urease by eliminating the binding of nickel irons to Comment: noncarbamylated protein [1]. Number of members: 20
Vif		Retroviral Vif (Viral infectivity) protein	Accession number: PF00559 Definition: Retroviral Vif (Viral infectivity) protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Swiss-Prot Gathering cutoffs: 25 25 Trusted cutoffs: 53.90 53.90 Noise cutoffs: 23.60 23.60 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95287525 Reference Title: Aberrant Gag protein composition of a human Reference Title: immunodeficiency virus type 1 vif mutant produced in Reference Title: primary lymphocytes. Reference Author: Simm M, Shahabuddin M, Chao W, Allan JS, Volsky DJ; Reference Location: J Virol 1995;69:4582-4586. Database Reference INTERPRO; IPR000475; Comment: -!- Human immunodeficiency virus type 1 (HIV-1) Vif is required for Comment: productive infection of T lymphocytes and macrophages. Virions Comment: produced in the absence of Vif have abnormal core morphology and Comment: those produced in primary T cells carry immature core proteins Comment: and low levels of mature capsid. Number of members: 503
Vpu		Vpu protein	Accession number: PF00558 Definition: Vpu protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Swiss-Prot Gathering cutoffs: 15 15 Trusted cutoffs: 15.50 15.50 Noise cutoffs: 13.60 13.60 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97479365 Reference Title: Enhancement of retroviral production from packaging cell Reference Title: lines expressing the human immunodeficiency type 1 VPU Reference Title: gene. Reference Author: Kobinger GP, Mouland AJ, Lalonde JP, Forget J, Cohen EA; Reference Location: Gene Ther 1997;4:868-874. Reference Number: [2] Reference Medline: 95156576 Reference Title: The human immunodeficiency virus type 1 Vpu protein Reference Title: specifically binds to the cytoplasmic domain of CD4: Reference Title: implications for the mechanism of degradation. Reference Author: Bour S, Schubert U, Strebel K; Reference Location: J Virol 1995;69:1510-1520.

Pfam	Prosite	Full Name	Description
			Reference Number: [3] Reference Medline: 97325981 Reference Title: Secondary structure and tertiary fold of the human Reference Title: immunodeficiency virus protein U (Vpu) cytoplasmic domain Reference Title: in solution. Reference Author: Willbold D, Hoffmann S, Rosch P; Reference Location: Eur J Biochem 1997;245:581-588. Database Reference: SCOP; 1vpu; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002094; Database Reference PDB; 1vpu ; 38; 81; Database reference: PFAMB; PB003303; Database reference: PFAMB; PB005882; Comment: -!- The Vpu protein contains an N-terminal transmembrane spanning region Comment: and a C-terminal cytoplasmic region. Comment: -!- The HIV-1 Vpu protein stimulates virus production by enhancing Comment: the release of viral particles from infected cells. Comment: -!- The VPU protein binds specifically to CD4. Number of members: 194
XPG_N	PDOC00658	XPG protein signatures	Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2]. XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets: - Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9]. - Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease. In addition to the proteins listed in the above groups, this family also includes: - Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs. - Yeast EXO1 (DHS1), a protein with probably the same function as exo1. - Yeast DIN7. Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it

Pfam	Prosite	Full Name	Description
			spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset. We have developed two signature patterns for these proteins. The first corresponds to the central part of the N-region, the second to part of the I- region and includes the putative catalytic core pentapeptide. Description of pattern(s) and/or profile(s) Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Clarkson S.G. clarkson@medecine.unige.ch Last update November 1997 / Patterns and text revised. References [1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994). [2] Scherly D., Nouspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993). [3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993). [4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S , Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994). [5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994). [6] Szankasi P., Smith G.R. Science 267:1166-1169(1995). [7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993). [8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).

Pfam	Prosite	Full Name	Description
			[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).
Y phosphatase	PDOC00323	Tyrosine specific protein phosphatases signature and profiles	Tyrosine specific protein phosphatases (EC 3 1 3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below: Soluble PTPases. - PTPN1 (PTP-1B). - PTPN2 (T-cell PTPase; TC-PTP). - PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1- like domain (see <PDOC00566>) and could act at junctions between the membrane and cytoskeleton. - PTPN5 (STEP). - PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup. - PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP). - PTPN8 (70Z-PEP). - PTPN9 (MEG2). - PTPN12 (PTP-G1; PTP-P19). - Yeast PTP1. - Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway. - Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis. - Fission yeast pyp3 which contributes to the dephosphorylation of cdc2. - Yeast CDC14 which may be involved in chromosome segregation. - Yersinia virulence plasmid PTPAses (gene yopH). - Autographa californica nuclear polyhedrosis virus 19 Kd PTPase. Dual specificity PTPases. - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185. - DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues. - DUSP3 (VHR). - DUSP4 (HVH2). - DUSP5 (HVH3). - DUSP6 (Pyst1; MKP-3). - DUSP7 (Pyst2; MKP-X). - Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3. - Yeast YVH1. - Vaccinia virus H1 PTPase; a dual specificity phosphatase. Receptor PTPases.

Pfam	Prosite	Full Name	Description

Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.

In the following table, the domain structure of known receptor PTPases is shown:

	Extracellular				Intracellular
	Ig	FN-3	CAH	MAM	PTPase
Leukocyte common antigen (LCA) (CD45)	0	2	0	0	2
Leukocyte antigen related (LAR)	3	8	0	0	2
Drosophila DLAR	3	9	0	0	2
Drosophila DPTP	2	2	0	0	2
PTP-alpha (LRP)	0	0	0	0	2
PTP-beta	0	16	0	0	1
PTP-gamma	0	1	1	0	2
PTP-delta	0	>7	0	0	2
PTP-epsilon	0	0	0	0	2
PTP-kappa	1	4	0	1	2
PTP-mu	1	4	0	1	2
PTP-zeta	0	1	1	0	2

PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.

We derived a signature pattern for PTPase domains centered on the active site cysteine.

There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.

Description of pattern(s) and/or profile(s)

Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue]
Sequences known to belong to this class detected by the pattern ALL, except for nine sequences.
Other sequence(s) detected in SWISS-PROT 3.

Sequences known to belong to this class detected by the 1st profile ALL.
Other sequence(s) detected in SWISS-PROT 2.

Sequences known to belong to this class detected by the 2nd

Pfam	Prosite	Full Name	Description
			profile ALL dual type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the 3rd profile ALL PTP type PTPases. Other sequence(s) detected in SWISS-PROT NONE. Note the M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases. Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. References [1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991). [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992). [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991). [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989). [5] Hunter T. Cell 58:1013-1016(1989).
Zein		Zein seed storage protein	Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Proteins 1993;15:88-99. Database Reference INTERPRO; IPR002530; Comment: Zeins are seed storage proteins. They are unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48
zf-AN1		AN1-like Zinc finger	Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16 Trusted cutoffs: 16.40 16.40

Pfam	Prosite	Full Name	Description
			Noise cutoffs: 7.30 7.30 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93292985 Reference Title: Two related localized mRNAs from Xenopus laevis encode Reference Title: ubiquitin-like fusion proteins. Reference Author: Linnen JM, Bailey CP, Weeks DL; Reference Location: Gene 1993;128:181-188. Database reference: SMART; ZnF_AN1; Database Reference INTERPRO; IPR000058; Comment: Zinc finger at the C-terminus of An1 Swiss:Q91889, a ubiquitin-like Comment: protein in Xenopus laevis. Comment: The following pattern describes the zinc finger. Comment: C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C Comment: Where X can be any amino acid, and numbers in brackets Comment: indicate the number of residues. Number of members: 18
zf-B_box	PDOC50015	B-box zinc finger	Accession number: PF00643 Definition: B-box zinc finger. Author: Bateman A Alignment method of seed: pftools Source of seed members: Prosite Gathering cutoffs: 25 25 Trusted cutoffs: 26.00 26.00 Noise cutoffs: 24.50 29.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Database Reference: SCOP; 1fre; fa; [SCOP-USA][CATH-PDBSUM] Database reference: PROSITE_PROFILE; PS50119; Database Reference: PROSITE; PDOC50015 Database Reference INTERPRO; IPR002991; Database Reference PDB; 1fre ; 4; 42; Database reference: PFAMB; PB002777; Database reference: PFAMB; PB010625; Database reference: PFAMB; PB041771; Number of members: 44
zf-CONSTANS		CONSTANS family zinc finger	Accession number. PF01760 Definition: CONSTANS family zinc finger Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1072 (release 4.2) Gathering cutoffs: 25 10 Trusted cutoffs: 76.10 17.20 Noise cutoffs: 9.70 9.70 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95211836 Reference Title: The CONSTANS gene of Arabidopsis promotes flowering and Reference Title: encodes a protein showing similarities to zinc finger Reference Title: transcription factors. Reference Author: Putterill J, Robson F, Lee K, Simon R, Coupland G; Reference Location: Cell 1995;80:847-857. Database Reference INTERPRO; IPR002926; Number of members: 45
zf-DHHC		DHHC zinc finger domain	Accession number: PF01529 Definition: DHHC zinc finger domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_945 (release 4.0) Gathering cutoffs: 22 22

Pfam	Prosite	Full Name	Description
			Trusted cutoffs: 22.40 22.40 Noise cutoffs: -22.40 -22.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99250263 Reference Title: The drosophila STAM gene homolog is in a tight gene Reference Title: cluster, and its expression correlates to that of the Reference Title. adjacent gene ial. Reference Author: Mesilaty-Gross S, Reich A, Motro B, Wides R; Reference Location: Gene 1999;231:173-186. Reference Number: [2] Reference Medline: 97315340 Reference Title: Variations of the C2H2 zinc finger motif in the yeast Reference Title: genome and classification of yeast zinc finger proteins. Reference Author: Bohm S, Frishman D, Mewes HW; Reference Location: Nucleic Acids Res 1997;25:2464-2469. Reference Number: [3] Reference Medline: 99321009 Reference Title: The DHHC domain: a new highly conserved cysteine-rich Reference Title: motif. Reference Author: Putilina T, Wong P, Gentleman S; Reference Location: Mol Cell Biochem 1999;195:219-226. Reference Number: [4] Reference Medline: 10490616 Reference Title: Erf2, a Novel Gene Product That Affects the Localization Reference Title: and Palmitoylation of Ras2 in Saccharomyces cerevisiae. Reference Author: Bartels DJ, Mitchell DA, Dong X, Deschenes RJ; Reference Location: Mol Cell Biol 1999;19:6775-6787. Database Reference INTERPRO; IPR001594; Comment: This domain is also known as NEW1 [2]. This domain is Comment: predicted to be a zinc binding domain. The function Comment: of this domain is unknown, but it has been predicted to Comment: be involved in protein-protein or protein-DNA Comment: interactions [3]. Number of members: 34
zf-MYND		MYND finger	Accession number: PF01753 Definition: MYND finger Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 11 11 Trusted cutoffs: 17.30 17.30 Noise cutoffs: 5.50 5.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96203118 Reference Title: DEAF-1, a novel protein that binds an essential region in a Reference Title: Deformed response element. Reference Author: Gross CT, McGinnis W; Reference Location: EMBO J 1996;15:1961-1970. Reference Number: [2] Reference Medline: 98079069 Reference Title: Molecular cloning, sequence analysis, expression, and Reference Title: tissue distribution of suppressin, a novel suppressor of Reference Title: cell cycle entry. Reference Author: LeBoeuf RD, Ban EM, Green MM, Stone

Pfam	Prosite	Full Name	Description
			AS, Propst SM, Blalock Reference Author: JE, Tauber JD; Reference Location: J Biol Chem 1998;273:361-368. Database Reference INTERPRO; IPR002893; Number of members: 48
Zn_carbOpept	PDOC00123	Zinc carboxypeptidases, zinc-binding regions signatures	There are a number of different types of zinc-dependent carboxypeptidases (EC 3.4.17.-) [1,2]. All these enzymes seem to be structurally and functionally related. The enzymes that belong to this family are listed below. - Carboxypeptidase A1 (EC 3 4.17.1), a pancreatic digestive enzyme that can removes all C-terminal amino acids with the exception of Arg, Lys and Pro. - Carboxypeptidase A2 (EC 3.4.17.15), a pancreatic digestive enzyme with a specificity similar to that of carboxypeptidase A1, but with a preference for bulkier C-terminal residues. - Carboxypeptidase B (EC 3.4.17.2), also a pancreatic digestive enzyme, but that preferentially removes C-terminal Arg and Lys. - Carboxypeptidase N (EC 3.4.17.3) (also known as arginine carboxypeptidase), a plasma enzyme which protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation. - Carboxypeptidase H (EC 3.4.17.10) (also known as enkephalin convertase or carboxypeptidase E), an enzyme located in secretory granules of pancreatic islets, adrenal gland, pituitary and brain. This enzyme removes residual C-terminal Arg or Lys remaining after initial endoprotease cleavage during prohormone processing. - Carboxypeptidase M (EC 3.4.17.12), a membrane bound Arg and Lys specific enzyme. It is ideally situated to act on peptide hormones at local tissue sites where it could control their activity before or after interaction with specific plasma membrane receptors. - Mast cell carboxypeptidase (EC 3.4.17.1), an enzyme with a specificity to carboxypeptidase A, but found in the secretory granules of mast cells. - Streptomyces griseus carboxypeptidase (Cpase SG) (EC 3.4.17.-) [3], which combines the specificities of mammalian carboxypeptidases A and B. - Thermoactinomyces vulgaris carboxypeptidase T (EC 3.4.17.18) (CPT) [4], which also combines the specificities of carboxypeptidases A and B. - AEBP1 [5], a transcriptional repressor active in preadipocytes. AEBP1 seems to regulate transcription by cleavage of other transcriptional proteins. - Yeast hypothetical protein YHR132c. All of these enzymes bind an atom of zinc. Three conserved residues are implicated in the binding of the zinc atom: two histidines and a glutamic acid We have derived two signature patterns which contain these three zinc-ligands.

Pfam	Prosite	Full Name	Description
			Description of pattern(s) and/or profile(s) Consensus pattern [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]- [STAG]-x(6)-[LIVMFYTA] [H and E are zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT Bacillus sphaericus endopeptidase I which hydrolyses the gamma-D-Glu-(L)meso-diaminopimelic acid bond of spore cortex peptidoglycan [6] and which is possibly distantly related to zinc carboxypeptidases. Consensus pattern H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW] [H is a zinc ligand] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 40. Note if a protein includes both signatures, the probability of it being a eukaryotic zinc carboxypeptidase is 100% Note these proteins belong to families M14A/M14B in the classification of peptidases [7,E1]. Last update November 1995 / Patterns and text revised. References [1] Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A. J. Biol. Chem. 264:13165-13170(1989). [2] Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F., Serafin W.E. J. Biol. Chem. 264:20094-20099(1989). [3] Narahashi Y. J. Biochem. 107:879-886(1990). [4] Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I., Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P., Galperina O.V., Matz M.V., Stepanov V.M. Eur. J. Biochem. 208:281-288(1992). [5] He G.-P., Muise A., Li A.W., Ro H.-S. Nature 378:92-96(1995). [6] Hourdou M.-L., Guinand M., Vacheron M.J., Michel G., Denoroy L., Duez C.M., Englebert S., Joris B., Weber G., Ghuysen J.-M. Biochem. J. 292:563-570(1993). [7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995). [E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt
ZZ		Zinc finger present in dystrophin, CBP/p300	Accession number. PF00569 Definition: Zinc finger present in dystrophin, CBP/p300 Author: SMART Alignment method of seed: Manual Source of seed members: Alignment kindly provided by SMART Gathering cutoffs: 14 14 Trusted cutoffs: 14.60 14.60 Noise cutoffs: 10.90 10.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96402609 Reference Title: ZZ and TAZ: new putative zinc fingers in

Pfam	Prosite	Full Name	Description
			dystrophin and Reference Title: other proteins. Reference Author: Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder Reference Author: SJ; Reference Location: Trends Biochem Sci 1996;21:11-13. Database Reference: EXPERT; Chris.Ponting@human-anatomy.oxford.ac.uk; Database Reference INTERPRO; IPR000433; Database reference: PFAMB; PB041629; Comment: ZZ in dystrophin binds calmodulin Comment: Putative zinc finger; binding not yet shown. Number of members: 87

AA. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically targeted to (1) a particular organelle or intracellular compartment, (2) interact with a particular molecule or (3) for secretion outside of a host cell. Example of polypeptides comprising signal peptides include, without limitation, secreted proteins, soluble proteins, receptors, proteins retained in the ER, etc.

These proteins comprising signal peptides are useful to modulate ligand-receptor interactions, cell-to-cell communication, signal transduction, intracellular communication, and activities and/or chemical cascades that take part in an organism outside or within of any particular cell.

One class of such proteins are soluble proteins which are transported out of the cell. These proteins can act as ligands that bind to receptor to trigger signal transduction or to permit communication between cells.

Another class is receptor proteins which also comprise a retention domain that lodges the receptor protein in the membrane when the cell transports the receptor to the surface of the cell. Like the soluble ligands, receptors can also modulate signal transduction and communication between cells.

In addition the signal peptide itself can serve as a ligand for some receptors. An example is the interaction of the ER targeting signal peptide with the signal recognition particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the resulting SRP complex then binds to docking proteins located on the surface of the ER, prompting transfer of the protein into the ER.

A description of signal peptide residue composition is described below in Subsection IV.C.1.

III. Methods of Modulating Polypeptide Production

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in Tables 1 and 2 of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation:

Suppression methods, such as

- Antisense
- Ribozymes
- Co-suppression
- Insertion of Sequences into the Gene to be Modulated
- Regulatory Sequence Modulation.

as well as Methods for Enhancing Production, such as

Insertion of Exogenous Sequences; and

Regulatory Sequence Modulation.

III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful, for instance, to tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437 (1991)) or to influence seed size (WO98/07842) or to provoke cell ablation (Mariani et al., Nature 357: 384-387 (1992).

As described above, a number of methods can be used to inhibit gene expression in plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell, insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the endogenous gene of interest, and the like.

III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*, e.g., Sheehy et al., *Proc. Nat. Acad. Sci.* USA, 85:8805 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA and down-regulate translation.

III.A.3. Co-Suppression

Another method of suppression is by introducing an exogenous copy of the gene to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to prevent the accumulation of mRNA. A detailed description of this method is described above.

III.A.4. Insertion of Sequences into the Gene to be Modulated

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* 26:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* 38:393 (1998), H. Albert et al., *Plant J.* 7:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

III.A.5. Regulatory SequenceModulation

The SDFs described in Tables 1 and 2, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

III.B.2. Regulatory Sequence Modulation

The SDFs of Tables 1 and 2, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all 11 promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in Tables 1 and 2, fragments thereof, and substantially similar sequences thereto.

Coding sequences for these proteins can be constructed as described above.

IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (*e.g.*, particle gun bombardment) as referenced below.

The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

(a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992); Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);

(b) **YAC:** Burke et al., Science 236:806-812 (1987);.

(c) **PAC:** Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);

(d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);

(e) **Lambda Phage Vectors**: Replacement Vector, e.g., Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g., Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL Press (1985);

(f) **T-DNA gene fusion vectors :**Walden et al., Mol Cell Biol 1: 175-194 (1990); and

(g) **Plasmid vectors:** Sambrook et al., infra.

Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimeraplast, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors of the invention can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational

initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide productionis desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF or the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species can modulate the expression of a native gene corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (*e.g.*, a copper inducible promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can

lead to co-suppression of the homologous endogeneous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (*e.g.*, seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue- specific manner can be altered.

IV. C Signal Peptides

SDFs of the present invention containing signal peptides are indicated in Tables 1 and 2. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vessicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of hydrophobic amino acids, such as those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the

chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in Tables 1 and 2 can be used to modulate polypeptide transport. Further variants of the native signal peptides described in Tables 1 and 2 are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. *See, e.g.* Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, Euphytica, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and

introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, *CM.*, *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, *AP.*, *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts,* pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium,Lupinus,*

Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, and, Zea.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The particular sequences of SDFs identified are provided in the attached Tables 1 and 2. One of ordinary skill in the art, having this data, can obtain cloned DNA fragments, synthetic DNA fragments or polypeptides constituting desired sequences by recombinant methodology known in the art or described herein.

EXAMPLES

The invention is illustrated by way of the following examples. The invention is not limited by these examples as the scope of the invention is defined solely by the claims following.

EXAMPLE 1: cDNA PREPARATION

A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* 15:707-720 (1998) High-efficiency cloning of Arabidopsis full-length

cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* 138:171 (1994) Oligo-capping a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and centrifugation. The polysomal RNA from the different tissues was pooled according to the following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in Tables 1 and 2 was poly(A)-containing polysomal mRNAs from inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the integrity of the polyA+ RNAs were evaluated.

Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in Tables 1 and 2 was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used

for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript™ using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated

oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

Following transformation, the libraries are ordered in microtiter plates and sequenced. The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as *E-coli* liba 010600" under the accession number **PTA-1161.**

EXAMPLE 2: SOUTHERN HYBRIDIZATIONS

The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percenatge of sequence identity between probe and target sequences that can be detected.

In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana.* As described above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

Buffers for nuclear DNA extraction

1. 10X HB

	1000 ml	
40 mM spermidine	10.2 g	Spermine (Sigma S-2876) and spermidine (Sigma S-2501)
10 mM spermine	3.5 g	Stabilize chromatin and the nuclear membrane
0.1 M EDTA (disodium)	37.2 g	EDTA inhibits nuclease
0.1 M Tris	12.1 g	Buffer
0.8 M KCl	59.6 g	Adjusts ionic strength for stability of nuclei

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)

Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

	<u>1000 ml</u>
N-lauroyl sarcosine (Sarkosyl)	20.0 g
0.1 M Tris	12.1 g
0.04 M EDTA (Disodium)	14.9 g

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100

80 ml Triton X-100

320 ml 1xHB (w/o β-ME and PMSF)

Prepare in advance; Triton takes some time to dissolve

A. Procedure

1. Prepare 1X H" buffer (keep ice-cold during use)

	1000 ml
10X HB	100 ml
2 M sucrose	250 ml a non-ionic osmoticum
Water	634 ml

Added just before use:

100 mM PMSF*	10 ml a protease inhibitor; protects nuclear membrane proteins
ß-mercaptoethanol	1 ml inactivates nuclease by reducing disulfide bonds

*100 mM PMSF

(phenyl methyl sulfonyl fluoride, Sigma P-7626)

(add 0.0875 g to 5 ml 100% ethanol)

2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.

3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

4. Filter the tissue suspension through several nylon filters into an ice-cold beaker. The first filtration is through a 250-micron membrane; the second is through an 85-micron membrane; the third is through a 50-micron membrane; and the fourth is through a 20-micron membrane. Use a large funnel to hold the filters. Filtration can be sped up by gently squeezing the liquid through the filters.

5. Centrifuge the filtrate at 1200 x g for 20 min. at 4°C to pellet the nuclei.

6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.

8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.

9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.

10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.

11. Add 20 μl of 10 mg/ml EtBr per ml of solution.

12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.

13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette. The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.

14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.

15. Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.

16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A260 concentration of the DNA.

17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known

and good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.

Protocol for Digestion of Genomic DNA

Protocol:

1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.

2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at –20°C for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.

3. Centrifuge the solution at 11,400 x g for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.

4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.

5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.

6. Set-up the lambda digestion-control for each DNA that you are digesting.

7. Incubate both the experimental and lambda digests overnight at 37°C. Spin down condensation in a microfuge before proceeding.

8. After digestion, add 2 μl of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load

in 1% TPE-agarose gel (TPE is 90 mM Tris-phosphate, 2 mM EDTA, pH 8). If the lambda DNA in the lambda control digests are completely digested, proceed with the precipitation of the genomic DNA in the digests.

9. Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in –20°C for at least 2 hours (preferably overnight).

EXCEPTION: *Arabidopsis* and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

10. Resuspend the DNA in an appropriate volume of TE (e.g., 22 µl x 50 blots = 1100 µl) and an appropriate volume of 10X loading dye (e.g., 2.4 µl x 50 blots = 120 µl). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3

Some guide points in digesting genomic DNA.

Species	**Genome Size**	**Size Relative to Arabidopsis**	**Genome Equivalent to 2 µg Arabidopsis DNA**	**Amount of DNA per blot**
Arabidopsis	120 Mb	1X	1X	2 µg
Brassica	1,100 Mb	9.2X	0.54X	10 µg
Corn	2,800 Mb	23.3X	0.43X	20 µg
Cotton	2,300 Mb	19.2X	0.52X	20 µg
Oat	11,300 Mb	94X	0.11X	20 µg
Rice	400 Mb	3.3X	0.75X	5 µg
Soybean	1,100 Mb	9.2X	0.54X	10 µg
Sugarbeet	758 Mb	6.3X	0.8X	10 µg
Sweetclover	1,100 Mb	9.2X	0.54X	10 µg
Wheat	16,000 Mb	133X	0.08X	20 µg

Yeast	15 Mb	0.12X	1X	0.25 μg

Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

1. For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.

2. Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.

3. The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.

4. A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)

5. The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.

6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.

B. <u>Protocol for PCR Amplification of Genomic Fragments in Arabidopsis</u>

<u>Amplification procedures</u>:

1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

Volume	Stock	Final Amount or Conc.
0.5 μl	~ 10 ng/μl genomic DNA[1]	5 ng
2.5 μl	**10X PCR buffer**	20 mM Tris, 50 mM KCl
0.75 μl	50 mM $MgCl_2$	1.5 mM
1 μl	10 pmol/μl Primer 1 (Forward)	10 pmol
1 μl	10 pmol/μl Primer 2 (Reverse)	10 pmol
0.5 μl	5 mM dNTPs	0.1 mM
0.1 μl	5 units/μl Platinum Taq™ (Life Technologies, Gaithersburg, MD) DNA Polymerase	1 units
(to 25 μl)	**Water**	

2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:

1) 94°C for 10 min. followed by

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
94 °C - 30 sec 62 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 58 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 53 °C - 30 sec 72 °C - 3 min

[1] Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

Quantification and Dilution of PCR Products:

1. The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100, 200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.

2. The amounts of PCR products can be estimated on the basis of the plasmid standard.

3. For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10-μl tip into the band while viewing though a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

C. Protocol for PCR-DIG-Labeling of DNA

Solutions:

Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM $MgCl_2$, 5 U/μl Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP **[1:5]**: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP **[1:10]**: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:15]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.875 mM dTTP, 0.125 mM DIG-11-dUTP)

TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min. and sterilize.

10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M $MgCl_2$ in autoclaved distilled water.

Procedure:

1. PCR reactions are performed in 25 μl volumes containing:

PCR buffer	1X
$MgCl_2$	1.5 mM
10X dNTP + DIG-11-dUTP	1X (please see the note below)
Platinum Taq™ Polymerase	1 unit
10 pg probe DNA	
10 pmol primer 1	

Note:	Use for:
10X dNTP + DIG-11-dUTP (1:5)	< 1 kb
10X dNTP + DIG-11-dUTP (1:10)	1 kb to 1.8 kb
10X dNTP + DIG-11-dUTP (1:15)	> 1.8 kb

2. The PCR reaction uses the following amplification cycles:

1) 94°C for 10 min.

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
95°C - 30 sec 61°C - 1 min 73°C - 5 min	95°C - 30 sec 59°C - 1 min 75°C - 5 min	95°C - 30 sec 51°C - 1 min 73°C - 5 min

5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).

3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.

4. The amount of DIG-labeled probe is determined as follows:

Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

DIG-labeled control DNA starting conc.	**Stepwise Dilution**	**Final Conc. (Dilution Name)**
5 ng/µl	1 µl in 49 µl TE	100 pg/µl (A)
100 pg/µl (A)	25 µl in 25 µl TE	50 pg/µl (B)
50 pg/µl (B)	25 µl in 25 µl TE	25 pg/µl (C)
25 pg/µl (C)	20 µl in 30 µl TE	10 pg/µl (D)

a. Serial deletions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.

b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.

c. The membrane is fixed by UV crosslinking.

d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.

e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.

f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

D. Prehybridization and Hybridization of Southern Blots

Solutions:

100% Formamide purchased from Gibco

20X SSC ($1X = 0.15$ M NaCl, 0.015 M Na_3citrate)

per L: 175 g NaCl

87.5 g Na_3citrate·$2H_2O$

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

Final Concentration	**Components**	**Volume (per 100 ml)**	**Stock**
50%	Formamide	50 ml	100%
5X	SSC	25 ml	20X
0.1%	Sarkosyl	0.5 ml	20%
0.02%	SDS	0.1 ml	20%
2%	Blocking Reagent	20 ml	10%
	Water	4.4 ml	

General Procedures:

1. Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm^2) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4–5 bags). Ensure that the bags are

lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.

3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.

4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.

5. Incubate with gentle agitation for at least 16 hours.

6. Proceed to medium stringency post-hybridization wash:

Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.

To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.

7. After the wash, proceed to immunological detection and CSPD development.

E. Procedure for Immunological Detection with CSPD

Solutions:

Buffer 1:	Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl; adjusted to pH 7.5 with NaoH)
Washing buffer:	Maleic acid buffer with 0.3% (v/v) Tween 20.

Blocking stock solution	10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.
Buffer 2 (1X blocking solution):	Dilute the stock solution 1:10 in Buffer 1.
Detection buffer:	0.1 M Tris, 0.1 M NaCl, pH 9.5

<u>Procedure</u>:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.

2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.

3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.

4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.

5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
6. The blots are equilibrated for 2–5 min in 60 ml detection buffer.

7. Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.

9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.

10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.

11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

Example 3: Transformation of Carrot Cells

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B_5 growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM $CaCl_2$ (B_5 -44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B_5-44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM $CaCl_2$ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution

containing 5 mM MES, 20 mM $CaCl_2$, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4 x 10^6 protoplasts per ml.

15-60 μg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Plt. Phys.* 79:988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures.*

An additional deposit of an *E. coli* Library, *E. coli*LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which encodes an amino acid sequence exhibiting at least 40% sequence identity to an amino acid sequence encoded by

(a) a nucleotide sequence described in Tables 1 and/or 2 or a fragment thereof; or

(b) a complement of a nucleotide sequence shown in Tables 1 and/or 2 or a fragment thereof.

2. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to

(a) a nucleotide sequence shown in Tables 1 and/or 2 or a fragment thereof; or

(b) a complement of a nucleotide sequence described in Tables 1 and/or 2 or a fragment thereof.

3. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to a gene comprising

(a) a nucleotide sequence shown in Tables 1 and/or 2 or a fragment thereof; or

(b) a complement of a nucleotide sequence described in Tables 1 and/or 2 or a fragment thereof.

4. An isolated nucleic acid molecule which is the reverse of the isolated nucleotide sequence according to claim 1, such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of said isolated nucleotide sequence according to claim 1.

5. An isolated nucleic acid molecule comprising a nucleic acid capable of hybridizing to a nucleic acid having a sequence selected from the group consisting of:

(a) a nucleotide sequence which is shown in Tables 1 and/or 2; and

(b) a nucleotide sequence which is complementary to a nucleotide sequence shown in Tables 1 and/or 2;

under conditions that permit formation of a nucleic acid duplex at a temperature from about 40°C and 48°C below the melting temperature of the nucleic acid duplex.

6. The nucleic acid molecule according to claim 1, wherein said nucleic acid comprises an open reading frame.

7. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid is capable of functioning as a promoter, a 3' end termination sequence, an untranslated region (UTR), or as a regulatory sequence.

8. The isolated nucleic acid molecule of claim 7, wherein said nucleic acid is a promoter and comprises a sequence selected from the group consisting of a TATA box sequence, a CAAT box sequence, a motif of GCAATCG or any transcription-factor binding sequence, and any combination thereof.

9. The isolated nucleic acid molecule of claim 7, wherein the nucleic acid sequence is a regulatory sequence which is capable of promoting seed-specific expression, embryo-specific expression, ovule-specific expression, tapetum-specific expression or root-specific expression of a sequence or any combination thereof.

10. A vector construct comprising a nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is heterologous to any element in said vector construct.

11. A vector construct comprising:

(a) a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation; and

(b) a second nucleic acid having the sequence of the isolated nucleic acid molecule according to claim 1;

wherein said first and second nucleic acids are operably linked and wherein said second nucleic acid is heterologous to any element in said vector construct.

12. The vector construct according to claim 11, wherein said first nucleic acid is native to said second nucleic acid.

13. The vector construct according to claim 11, wherein said first nucleic acid is heterologous to said second nucleic acid.

14. A vector construct comprising:

(c) a first nucleic acid having the sequence of the isolated nucleic acid molecule according to claim 7; and

(d) a second nucleic acid;

wherein said first and second nucleic acids are operably linked and wherein said first nucleic acid is heterologous to any element in said vector construct.

15. The vector construct according to claim 14, wherein said first nucleic acid is native to said second nucleic acid.

16. The vector construct according to claim 14, wherein said first nucleic acid is heterologous to said second nucleic acid.

17. A host cell comprising an isolated nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is flanked by exogenous sequence.

18. A host cell comprising a vector construct of claim 10.

19. A host cell comprising a vector construct of claim 11.

20. A host cell comprising a vector construct of claim 12.

21. A host cell comprising a vector construct of claim 13.

22. A host cell comprising a vector construct of claim 14.

23. A host cell comprising a vector construct of claim 15.

24. A host cell comprising a vector construct of claim 16.

25. An isolated polypeptide comprising an amino acid sequence

(a) exhibiting at least 40% sequence identity of an amino acid sequence encoded by a sequence shown in Tables 1 and/or 2 or a fragment thereof; and

(b) capable of exhibiting at least one of the biological activities of the polypeptide encoded by said nucleotide sequence shown in Tables 1 and/or 2 or a fragment thereof.

26. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 75% sequence identity to an amino acid sequence encoded by a sequence shown in Tables 1 and/or 2 or a fragment thereof.

27. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 85% sequence identity to an amino acid sequence encoded by a sequence shown in Tables 1 and/or 2 or a fragment thereof.

28. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 90% sequence identity to an amino acid sequence encoded by a sequence shown in Tables 1 and/or 2 or a fragment thereof.

29. An antibody capable of binding the isolated polypeptide of claim 25.

30. A method of introducing an isolated nucleic acid into a host cell comprising:

(a) providing an isolated nucleic acid molecule according to claim 1; and

(b) contacting said isolated nucleic with said host cell under conditions that permit insertion of said nucleic acid into said host cell.

31. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 10.

32. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 11.

33. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 12.

34. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 13.

35. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 14.

36. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 15.

37. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 16.

38. A method of modulating transcription and/or translation of a nucleic acid in a host cell comprising:

(a) providing the host cell of claim 17; and

(b) culturing said host cell under conditions that permit transcription or translation.

39. A method for detecting a nucleic acid in a sample which comprises:

(a) providing an isolated nucleic acid molecule according to claim 1;

(b) contacting said isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of said isolated nucleic acid molecule with the sequence of DNA in said sample; and

(c) analyzing the result of said comparison.

40. The method according to claim 39, wherein said isolated nucleic acid molecule and said sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.

41. A plant or cell of a plant which comprises a nucleic acid molecule according to claim 1 which is exogenous to said plant or plant cell.

42. A plant or cell of a plant which comprises a nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is heterologous to said plant or said cell of a plant.

43. A plant or cell of a plant which has been transformed with a nucleic acid molecule according to claim 1.

44. A plant or cell of a plant which comprises a vector construct according to claim 10.

45. A plant or cell of a plant which has been transformed with a vector construct according to claim 10.

46. A plant which has been regenerated from a plant cell according to claim 41.

47. A plant which has been regenerated from a plant cell according to claim 42.

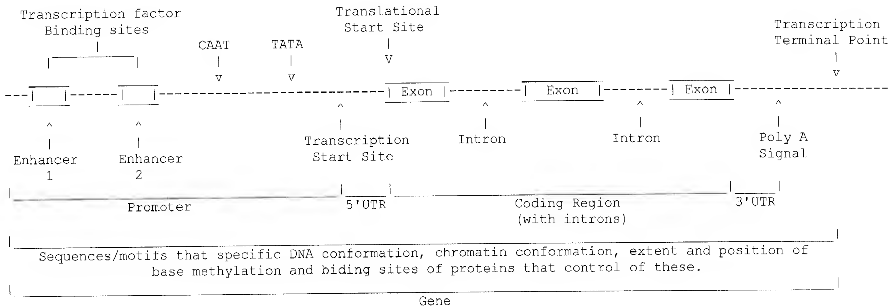
48. A plant which has been regenerated from a plant cell according to claim 43.

49. A plant which has been regenerated from a plant cell according to claim 44.

50. A plant which has been regenerated from a plant cell according to claim 45.

SCHEMATIC 1

SCHEMATIC OF A GENE



ABSTRACT OF THE DISCLOSURE

The present invention provides DNA molecules that constitute fragments of the genome of a plant, and polypeptides encoded thereby. The DNA molecules are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome, in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait.

POWER OF ATTORNEY

CERES, INC.
3007 Malibu Canyon Road
Malibu, CA 90265

I, Richard Hamilton, Chief Financial Officer of CERES, INC. of 3007 Malibu Canyon Road, Malibu, California 90265, grant Power of Attorney and authority to empower the following attorneys to act on behalf of CERES, INC. for executing Verified Statements (Declarations) Claiming Small Entity Status to be submitted to the U.S. Patent and Trademark Office in connection with the filing of provisional or regular patent applications on behalf of CERES, INC.

Raymond C. Stewart (Reg. No. 21,066)
Joseph A. Kolasch (Reg. No. 22,463)
Leonard R. Svensson (Reg. No. 30,330)
Gerald M. Murphy, Jr. (Reg. No.28,977)
Mark J. Nuell (Reg. No. 36,623)

This Power of Attorney is to remain in full force and effect until terminated by an official of CERES, INC.

By ____________________
Richard Hamilton

Date 9/24/98

Maximum Length Sequence:
related to:
Clone IDs:
39998
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1
- Ceres seq_id 1565539

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2
- Ceres seq_id 1565540
- Location of start within SEQ ID NO 1: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1
- gi No. 399298
- Description: STRESS-INDUCED KIN2 PROTEIN (COLD-INDUCED COR6.6 PROTEIN) >gi|1084343|pir||S22529 cold-regulated protein kin2 - Arabidopsis thaliana >gi|16230|emb|CAA38894| (X55053) cold regulated [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 54
- Location of Alignment in SEQ ID NO 2: from 1 to 54

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3
- Ceres seq_id 1565541
- Location of start within SEQ ID NO 1: at 252 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1152
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4
- Ceres seq_id 1565546

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 5
- Ceres seq_id 1565547
- Location of start within SEQ ID NO 4: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 6
- Ceres seq_id 1565548
- Location of start within SEQ ID NO 4: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Macrophage migration inhibitory factor (MIF)
- Location within SEQ ID NO 6: from 2 to 54 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 7
- Ceres seq_id 1565549
- Location of start within SEQ ID NO 4: at 473 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39126
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 8
- Ceres seq_id 1565556
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 9
- Ceres seq_id 1565557
- Location of start within SEQ ID NO 8: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 2
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 9: from 75 to 355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 10
- Ceres seq_id 1565558
- Location of start within SEQ ID NO 8: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 3
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 10: from 61 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 11
- Ceres seq_id 1565559
- Location of start within SEQ ID NO 8: at 195 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 4
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 11: from 39 to 319

Maximum Length Sequence:
related to:
Clone IDs:
39141
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 12
- Ceres seq_id 1565568
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 13
- Ceres seq_id 1565569
- Location of start within SEQ ID NO 12: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cys/Met metabolism PLP-dependent enzyme
- Location within SEQ ID NO 13: from 50 to 270 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 14
- Ceres seq_id 1565570
- Location of start within SEQ ID NO 12: at 240 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cys/Met metabolism PLP-dependent enzyme
- Location within SEQ ID NO 14: from 1 to 209 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 15
- Ceres seq_id 1565571
- Location of start within SEQ ID NO 12: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cys/Met metabolism PLP-dependent enzyme
- Location within SEQ ID NO 15: from 1 to 198 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
7008
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 16
- Ceres seq_id 1565576
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 17
- Ceres seq_id 1565577
- Location of start within SEQ ID NO 16: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 17: from 81 to 109 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 5
- gi No. 115484
- Description: CALMODULIN 1 >gi|71684|pir||MCPZDC calmodulin - carrot >gi|478632|pir||S22971 calmodulin - trumpet lily >gi|541839|pir||S40301 calmodulin - Red bryony >gi|2129970|pir||S70768 calmodulin CAM81 - garden petunia
- % Identity: 100
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 17: from 34 to 182

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 18
- Ceres seq_id 1565578
- Location of start within SEQ ID NO 16: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 18: from 48 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 6
- gi No. 115484
- Description: CALMODULIN 1 >gi|71684|pir||MCPZDC calmodulin - carrot >gi|478632|pir||S22971 calmodulin - trumpet lily >gi|541839|pir||S40301 calmodulin - Red bryony >gi|2129970|pir||S70768 calmodulin CAM81 - garden petunia
- % Identity: 100
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 18: from 1 to 149

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 19
- Ceres seq_id 1565579
- Location of start within SEQ ID NO 16: at 208 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 19: from 12 to 40 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 7
- gi No. 115484
- Description: CALMODULIN 1 >gi|71684|pir||MCPZDC calmodulin - carrot >gi|478632|pir||S22971 calmodulin - trumpet lily >gi|541839|pir||S40301 calmodulin - Red bryony >gi|2129970|pir||S70768 calmodulin CAM81 - garden petunia
- % Identity: 100
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 19: from 1 to 113

Maximum Length Sequence:
related to:
Clone IDs:
12548
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 20
- Ceres seq_id 1565612

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 21
- Ceres seq_id 1565613
- Location of start within SEQ ID NO 20: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enol-ase
- Location within SEQ ID NO 21: from 27 to 349 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 8
- gi No. 119350
- Description: ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >gi|81608|pir||JQ1187 phosphopyruvate hydratase (EC 4.2.1.11) - Arabidopsis thaliana >gi|16271|emb|CAA41114| (X58107) enolase [Arabidopsis thaliana] thaliana]
- % Identity: 99.4
- Alignment Length: 324
- Location of Alignment in SEQ ID NO 21: from 26 to 349

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 22
- Ceres seq_id 1565614
- Location of start within SEQ ID NO 20: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enol-ase
- Location within SEQ ID NO 22: from 2 to 324 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 9
- gi No. 119350
- Description: ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >gi|81608|pir||JQ1187 phosphopyruvate hydratase (EC 4.2.1.11) - Arabidopsis thaliana >gi|16271|emb|CAA41114| (X58107) enolase [Arabidopsis thaliana] thaliana]
- % Identity: 99.4
- Alignment Length: 324
- Location of Alignment in SEQ ID NO 22: from 1 to 324

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 23
- Ceres seq_id 1565615
- Location of start within SEQ ID NO 20: at 354 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enol-ase
- Location within SEQ ID NO 23: from 1 to 232 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 10
- gi No. 119350
- Description: ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >gi|81608|pir||JQ1187 phosphopyruvate hydratase (EC 4.2.1.11) - Arabidopsis thaliana >gi|16271|emb|CAA41114| (X58107) enolase [Arabidopsis thaliana] thaliana]
- % Identity: 99.4
- Alignment Length: 324
- Location of Alignment in SEQ ID NO 23: from 1 to 232

Maximum Length Sequence:
related to:
Clone IDs:
12638
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 24
- Ceres seq_id 1565616
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 25
- Ceres seq_id 1565617
- Location of start within SEQ ID NO 24: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxylesterases
- Location within SEQ ID NO 25: from 146 to 242 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 26
- Ceres seq_id 1565618
- Location of start within SEQ ID NO 24: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxylesterases
- Location within SEQ ID NO 26: from 126 to 222 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32346
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 27
- Ceres seq_id 1565633
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 28
- Ceres seq_id 1565634
- Location of start within SEQ ID NO 27: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 28: from 41 to 314 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 11
- gi No. 4490309
- Description: (AL035678) peroxidase ATP17a-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 314
- Location of Alignment in SEQ ID NO 28: from 1 to 314

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 29
- Ceres seq_id 1565635

- Location of start within SEQ ID NO 27: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 29: from 33 to 306 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 12
- gi No. 4490309
- Description: (AL035678) peroxidase ATP17a-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 314
- Location of Alignment in SEQ ID NO 29: from 1 to 306

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 30
- Ceres seq_id 1565636
- Location of start within SEQ ID NO 27: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 30: from 29 to 302 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 13
- gi No. 4490309
- Description: (AL035678) peroxidase ATP17a-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 314
- Location of Alignment in SEQ ID NO 30: from 1 to 302

Maximum Length Sequence:
related to:
Clone IDs:
22084
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 31
- Ceres seq_id 1565643

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 32
- Ceres seq_id 1565644
- Location of start within SEQ ID NO 31: at 214 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 14
- gi No. 2245029
- Description: (Z97341) limonene cyclase like protein [Arabidopsis thaliana]
- % Identity: 90.1
- Alignment Length: 263
- Location of Alignment in SEQ ID NO 32: from 24 to 284

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 33
- Ceres seq_id 1565645

- Location of start within SEQ ID NO 31: at 331 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 15
- gi No. 2245029
- Description: (Z97341) limonene cyclase like protein [Arabidopsis thaliana]
- % Identity: 90.1
- Alignment Length: 263
- Location of Alignment in SEQ ID NO 33: from 1 to 245

Maximum Length Sequence:
related to:
Clone IDs:
17690

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 34
- Ceres seq_id 1565646

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 35
- Ceres seq_id 1565647
- Location of start within SEQ ID NO 34: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic-type carbonic anhydrase
- Location within SEQ ID NO 35: from 109 to 339 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 16
- gi No. 1620077
- Description: (U42580) Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]
- % Identity: 72.3
- Alignment Length: 83
- Location of Alignment in SEQ ID NO 35: from 24 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 36
- Ceres seq_id 1565648
- Location of start within SEQ ID NO 34: at 457 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic-type carbonic anhydrase
- Location within SEQ ID NO 36: from 1 to 209 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 37
- Ceres seq_id 1565649
- Location of start within SEQ ID NO 34: at 475 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic-type carbonic anhydrase
- Location within SEQ ID NO 37: from 1 to 203 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22613
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 38
- Ceres seq_id 1565650
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 39
- Ceres seq_id 1565651
- Location of start within SEQ ID NO 38: at 231 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 39: from 79 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 17
- gi No. 3219271
- Description: (AB015315) MAP kinase kinase 4 [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 39: from 1 to 137

Maximum Length Sequence:
related to:
Clone IDs:
15980
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 40
- Ceres seq_id 1565658
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 41
- Ceres seq_id 1565659
- Location of start within SEQ ID NO 40: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 18
- gi No. 1931645
- Description: (U95973) Fe(II) transporter isolog [Arabidopsis thaliana]
- % Identity: 80.5
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 41: from 65 to 453

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 42
- Ceres seq_id 1565660
- Location of start within SEQ ID NO 40: at 195 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 19

- gi No. 1931645
- Description: (U95973) Fe(II) transporter isolog [Arabidopsis thaliana]
- % Identity: 80.5
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 42: from 1 to 389

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 43
- Ceres seq_id 1565661
- Location of start within SEQ ID NO 40: at 435 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 20
- gi No. 1931645
- Description: (U95973) Fe(II) transporter isolog [Arabidopsis thaliana]
- % Identity: 80.5
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 43: from 1 to 309

Maximum Length Sequence:
related to:
Clone IDs:
16846

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 44
- Ceres seq_id 1565662

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 45
- Ceres seq_id 1565663
- Location of start within SEQ ID NO 44: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S-adenosyl-L-homocysteine hydrolase
- Location within SEQ ID NO 45: from 58 to 448 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 21
- gi No. 2244750
- Description: (Z97335) adenosylhomocysteinase [Arabidopsis thaliana] >gi|3088579 (AF059581) S-adenosyl-L-homocysteine hydrolase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 45: from 23 to 487

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 46
- Ceres seq_id 1565664
- Location of start within SEQ ID NO 44: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S-adenosyl-L-homocysteine hydrolase
- Location within SEQ ID NO 46: from 36 to 426 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 22
- gi No. 2244750
- Description: (Z97335) adenosylhomocysteinase [Arabidopsis thaliana] >gi|3088579 (AF059581) S-adenosyl-L-homocysteine hydrolase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 46: from 1 to 465

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 47
- Ceres seq_id 1565665
- Location of start within SEQ ID NO 44: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S-adenosyl-L-homocysteine hydrolase
- Location within SEQ ID NO 47: from 18 to 408 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 23
- gi No. 2244750
- Description: (Z97335) adenosylhomocysteinase [Arabidopsis thaliana] >gi|3088579 (AF059581) S-adenosyl-L-homocysteine hydrolase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 47: from 1 to 447

Maximum Length Sequence:
related to:
Clone IDs:
18738
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 48
- Ceres seq_id 1565666

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 49
- Ceres seq_id 1565667
- Location of start within SEQ ID NO 48: at 220 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 49: from 71 to 212 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 50
- Ceres seq_id 1565668
- Location of start within SEQ ID NO 48: at 616 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 50: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 51
- Ceres seq_id 1565669
- Location of start within SEQ ID NO 48: at 679 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 51: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
19157
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 52
- Ceres seq_id 1565672
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 53
- Ceres seq_id 1565673
- Location of start within SEQ ID NO 52: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 53: from 107 to 172 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 54
- Ceres seq_id 1565674
- Location of start within SEQ ID NO 52: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 54: from 71 to 136 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24175
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 55
- Ceres seq_id 1565683
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 56
- Ceres seq_id 1565684
- Location of start within SEQ ID NO 55: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 24
- gi No. 3123264
- Description: 60S RIBOSOMAL PROTEIN L27
>gi|2244857|emb|CAB10279.1| (Z97337) ribosomal protein [Arabidopsis thaliana]

- % Identity: 100
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 56: from 1 to 89

Maximum Length Sequence:
related to:
Clone IDs:
28637
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 57
- Ceres seq_id 1565688
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 58
- Ceres seq_id 1565689
- Location of start within SEQ ID NO 57: at 845 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 25
- gi No. 2252827
- Description: (AF013293) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 58: from 9 to 297

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 59
- Ceres seq_id 1565690
- Location of start within SEQ ID NO 57: at 869 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 26
- gi No. 2252827
- Description: (AF013293) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 59: from 1 to 289

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 60
- Ceres seq_id 1565691
- Location of start within SEQ ID NO 57: at 1079 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 27
- gi No. 2252827
- Description: (AF013293) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 60: from 1 to 219

Maximum Length Sequence:
related to:

Clone IDs:
31027
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 61
- Ceres seq_id 1565692
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 62
- Ceres seq_id 1565693
- Location of start within SEQ ID NO 61: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 28
- gi No. 2252832
- Description: (AF013293) contains similarity to S. cerevisiae ADR1 gene (PID:g924931) [Arabidopsis thaliana]
- % Identity: 97.6
- Alignment Length: 253
- Location of Alignment in SEQ ID NO 62: from 1 to 247

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 63
- Ceres seq_id 1565694
- Location of start within SEQ ID NO 61: at 299 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 29
- gi No. 2252832
- Description: (AF013293) contains similarity to S. cerevisiae ADR1 gene (PID:g924931) [Arabidopsis thaliana]
- % Identity: 97.6
- Alignment Length: 253
- Location of Alignment in SEQ ID NO 63: from 1 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 64
- Ceres seq_id 1565695
- Location of start within SEQ ID NO 61: at 329 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 30
- gi No. 2252832
- Description: (AF013293) contains similarity to S. cerevisiae ADR1 gene (PID:g924931) [Arabidopsis thaliana]
- % Identity: 97.6
- Alignment Length: 253
- Location of Alignment in SEQ ID NO 64: from 1 to 138

Maximum Length Sequence:
related to:
Clone IDs:
35408
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 65
- Ceres seq_id 1565723
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 66
- Ceres seq_id 1565724
- Location of start within SEQ ID NO 65: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 31
- gi No. 3935185
- Description: (AC004557) F17L21.28 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 432
- Location of Alignment in SEQ ID NO 66: from 1 to 432

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 67
- Ceres seq_id 1565725
- Location of start within SEQ ID NO 65: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 32
- gi No. 3935185
- Description: (AC004557) F17L21.28 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 432
- Location of Alignment in SEQ ID NO 67: from 1 to 413

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 68
- Ceres seq_id 1565726
- Location of start within SEQ ID NO 65: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 33
- gi No. 3935185
- Description: (AC004557) F17L21.28 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 432
- Location of Alignment in SEQ ID NO 68: from 1 to 410

Maximum Length Sequence:
related to:
Clone IDs:
38527

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 69
- Ceres seq_id 1565746

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 70
- Ceres seq_id 1565747
- Location of start within SEQ ID NO 69: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 70: from 11 to 346 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 34
- gi No. 3882355
- Description: (U92460) 12-oxophytodienoate reductase OPR1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 70: from 1 to 372

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 71
- Ceres seq_id 1565748
- Location of start within SEQ ID NO 69: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 71: from 1 to 329 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 35
- gi No. 3882355
- Description: (U92460) 12-oxophytodienoate reductase OPR1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 71: from 1 to 355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 72
- Ceres seq_id 1565749
- Location of start within SEQ ID NO 69: at 470 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 72: from 1 to 213 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 36
- gi No. 3882355
- Description: (U92460) 12-oxophytodienoate reductase OPR1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 72: from 1 to 239

Maximum Length Sequence:
related to:
Clone IDs:
11370
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 73
- Ceres seq_id 1565769

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 74
- Ceres seq_id 1565770
- Location of start within SEQ ID NO 73: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 74: from 26 to 103 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 75
- Ceres seq_id 1565771
- Location of start within SEQ ID NO 73: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 75: from 5 to 82 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 76
- Ceres seq_id 1565772
- Location of start within SEQ ID NO 73: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 76: from 4 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
9756
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 77
- Ceres seq_id 1565777

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 78
- Ceres seq_id 1565778
- Location of start within SEQ ID NO 77: at 249 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 37
- gi No. 4204289
- Description: (AC003027) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 78: from 1 to 152

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 79
- Ceres seq_id 1565779
- Location of start within SEQ ID NO 77: at 264 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 38

- gi No. 4204289
- Description: (AC003027) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 79: from 1 to 147

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 80
- Ceres seq_id 1565780
- Location of start within SEQ ID NO 77: at 462 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 39
- gi No. 4204289
- Description: (AC003027) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 80: from 1 to 81

Maximum Length Sequence:
related to:
Clone IDs:
13012

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 81
- Ceres seq_id 1565787

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 82
- Ceres seq_id 1565788
- Location of start within SEQ ID NO 81: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 40
- gi No. 445588
- Description: Pch313 protein [Prunus persica]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 82: from 61 to 71

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 83
- Ceres seq_id 1565789
- Location of start within SEQ ID NO 81: at 732 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 83: from 1 to 128 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 84
- Ceres seq_id 1565790
- Location of start within SEQ ID NO 81: at 891 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 84: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22158
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 85
- Ceres seq_id 1565799
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 86
- Ceres seq_id 1565800
- Location of start within SEQ ID NO 85: at 112 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRP54-type protein
- Location within SEQ ID NO 86: from 28 to 295 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 41
- gi No. 586036
- Description: SIGNAL RECOGNITION PARTICLE 54 KD PROTEIN 1 (SRP54) >gi|629560|pir||S42550 signal recognition particle 54K protein - Arabidopsis thaliana >gi|304111 (L19997) signal recognition particle 54 kDa subunit [Arabidopsis thaliana] gene.
- % Identity: 100
- Alignment Length: 479
- Location of Alignment in SEQ ID NO 86: from 1 to 479

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 87
- Ceres seq_id 1565801
- Location of start within SEQ ID NO 85: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRP54-type protein
- Location within SEQ ID NO 87: from 12 to 279 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 42
- gi No. 586036
- Description: SIGNAL RECOGNITION PARTICLE 54 KD PROTEIN 1 (SRP54) >gi|629560|pir||S42550 signal recognition particle 54K protein - Arabidopsis thaliana >gi|304111 (L19997) signal recognition particle 54 kDa subunit [Arabidopsis thaliana] gene.
- % Identity: 100
- Alignment Length: 479
- Location of Alignment in SEQ ID NO 87: from 1 to 463

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 88
- Ceres seq_id 1565802
- Location of start within SEQ ID NO 85: at 265 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRP54-type protein
- Location within SEQ ID NO 88: from 1 to 244 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 43
- gi No. 586036
- Description: SIGNAL RECOGNITION PARTICLE 54 KD PROTEIN 1 (SRP54) >gi|629560|pir||S42550 signal recognition particle 54K protein - Arabidopsis thaliana >gi|304111 (L19997) signal recognition particle 54 kDa subunit [Arabidopsis thaliana] gene.
- % Identity: 100
- Alignment Length: 479
- Location of Alignment in SEQ ID NO 88: from 1 to 428

Maximum Length Sequence:
related to:
Clone IDs:
20342
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 89
- Ceres seq_id 1565810

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 90
- Ceres seq_id 1565811
- Location of start within SEQ ID NO 89: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 44
- gi No. 452470
- Description: (U05218) ATP sulfurylase [Arabidopsis thaliana]
- % Identity: 97.2
- Alignment Length: 463
- Location of Alignment in SEQ ID NO 90: from 1 to 463

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 91
- Ceres seq_id 1565812
- Location of start within SEQ ID NO 89: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 45
- gi No. 452470
- Description: (U05218) ATP sulfurylase [Arabidopsis thaliana]
- % Identity: 97.2
- Alignment Length: 463
- Location of Alignment in SEQ ID NO 91: from 1 to 460

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 92
- Ceres seq_id 1565813
- Location of start within SEQ ID NO 89: at 342 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 46
- gi No. 452470
- Description: (U05218) ATP sulfurylase [Arabidopsis thaliana]
- % Identity: 97.2
- Alignment Length: 463
- Location of Alignment in SEQ ID NO 92: from 1 to 373

Maximum Length Sequence:
related to:
Clone IDs:
16740
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 93
- Ceres seq_id 1565816

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 94
- Ceres seq_id 1565817
- Location of start within SEQ ID NO 93: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 94: from 68 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 95
- Ceres seq_id 1565818
- Location of start within SEQ ID NO 93: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 95: from 10 to 68 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 96
- Ceres seq_id 1565819
- Location of start within SEQ ID NO 93: at 250 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 96: from 9 to 67 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
18724
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 97
- Ceres seq_id 1565820

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 98
- Ceres seq_id 1565821

- Location of start within SEQ ID NO 97: at 308 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 47
- gi No. 4587515
- Description: (AC007060) EST gb|Z37678 comes from this gene. [Arabidopsis thaliana]
- % Identity: 95.2
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 98: from 1 to 305

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 99
- Ceres seq_id 1565822
- Location of start within SEQ ID NO 97: at 353 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 48
- gi No. 4587515
- Description: (AC007060) EST gb|Z37678 comes from this gene. [Arabidopsis thaliana]
- % Identity: 95.2
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 99: from 1 to 290

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 100
- Ceres seq_id 1565823
- Location of start within SEQ ID NO 97: at 461 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 49
- gi No. 4587515
- Description: (AC007060) EST gb|Z37678 comes from this gene. [Arabidopsis thaliana]
- % Identity: 95.2
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 100: from 1 to 254

Maximum Length Sequence:
related to:
Clone IDs:
15529
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 101
- Ceres seq_id 1565828

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 102
- Ceres seq_id 1565829
- Location of start within SEQ ID NO 101: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 102: from 5 to 405 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 50
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 102: from 1 to 449

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 103
- Ceres seq_id 1565830
- Location of start within SEQ ID NO 101: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 103: from 1 to 357 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 51
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 103: from 1 to 401

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 104
- Ceres seq_id 1565831
- Location of start within SEQ ID NO 101: at 410 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 104: from 1 to 304 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 52
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 104: from 1 to 348

Maximum Length Sequence:
related to:
Clone IDs:
29408
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 105
- Ceres seq_id 1565836

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 106
- Ceres seq_id 1565837
- Location of start within SEQ ID NO 105: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 53
- gi No. 4769004
- Description: (AF140598) ring-box protein 1 [Homo sapiens] >gi|4769006|gb|AAD29716.1|AF140599_1 (AF140599) ring-box protein 1 [Mus musculus] >gi|4809216|gb|AAD30146.1|AF142059_1 (AF142059) RING finger protein [Homo sapiens]
- % Identity: 88.8
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 106: from 21 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 107
- Ceres seq_id 1565838
- Location of start within SEQ ID NO 105: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 54
- gi No. 4769004
- Description: (AF140598) ring-box protein 1 [Homo sapiens] >gi|4769006|gb|AAD29716.1|AF140599_1 (AF140599) ring-box protein 1 [Mus musculus] >gi|4809216|gb|AAD30146.1|AF142059_1 (AF142059) RING finger protein [Homo sapiens]
- % Identity: 88.8
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 107: from 12 to 109

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 108
- Ceres seq_id 1565839
- Location of start within SEQ ID NO 105: at 437 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
27439
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 109
- Ceres seq_id 1565840

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 110
- Ceres seq_id 1565841
- Location of start within SEQ ID NO 109: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 110: from 1 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 111
- Ceres seq_id 1565842
- Location of start within SEQ ID NO 109: at 288 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 111: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 112
- Ceres seq_id 1565843
- Location of start within SEQ ID NO 109: at 357 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 112: from 123 to 281 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
27467
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 113
- Ceres seq_id 1565863
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 114
- Ceres seq_id 1565864
- Location of start within SEQ ID NO 113: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nucleosome assembly protein (NAP)
- Location within SEQ ID NO 114: from 26 to 219 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
41424
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 115
- Ceres seq_id 1565871
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 116
- Ceres seq_id 1565872
- Location of start within SEQ ID NO 115: at 155 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Serine carboxypeptidase

- Location within SEQ ID NO 116: from 30 to 196 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
38429
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 117
- Ceres seq_id 1565880
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 118
- Ceres seq_id 1565881
- Location of start within SEQ ID NO 117: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 118: from 36 to 443 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 55
- gi No. 4836883
- Description: (AC007260) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 370
- Location of Alignment in SEQ ID NO 118: from 1 to 370

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 119
- Ceres seq_id 1565882
- Location of start within SEQ ID NO 117: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 119: from 28 to 435 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 56
- gi No. 4836883
- Description: (AC007260) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 370
- Location of Alignment in SEQ ID NO 119: from 1 to 362

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 120
- Ceres seq_id 1565883
- Location of start within SEQ ID NO 117: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 120: from 26 to 433 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 57

- gi No. 4836883
- Description: (AC007260) lcl|prt_seq No definition line found [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 370
- Location of Alignment in SEQ ID NO 120: from 1 to 360

Maximum Length Sequence:
related to:
Clone IDs:
12880
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 121
- Ceres seq_id 1565898

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 122
- Ceres seq_id 1565899
- Location of start within SEQ ID NO 121: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin central region
- Location within SEQ ID NO 122: from 224 to 501 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 58
- gi No. 3341679
- Description: (AC003672) dynamin-like protein phragmoplastin 12 [Arabidopsis thaliana]
- % Identity: 73.5
- Alignment Length: 614
- Location of Alignment in SEQ ID NO 122: from 4 to 613

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 123
- Ceres seq_id 1565900
- Location of start within SEQ ID NO 121: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin central region
- Location within SEQ ID NO 123: from 221 to 498 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 59
- gi No. 3341679
- Description: (AC003672) dynamin-like protein phragmoplastin 12 [Arabidopsis thaliana]
- % Identity: 73.5
- Alignment Length: 614
- Location of Alignment in SEQ ID NO 123: from 1 to 610

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 124
- Ceres seq_id 1565901
- Location of start within SEQ ID NO 121: at 258 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin central region
- Location within SEQ ID NO 124: from 196 to 473 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 60
- gi No. 3341679
- Description: (AC003672) dynamin-like protein phragmoplastin 12 [Arabidopsis thaliana]
- % Identity: 73.5
- Alignment Length: 614
- Location of Alignment in SEQ ID NO 124: from 1 to 585

Maximum Length Sequence:
related to:
Clone IDs:
30989
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 125
- Ceres seq_id 1565922
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 126
- Ceres seq_id 1565923
- Location of start within SEQ ID NO 125: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 126: from 108 to 265 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 127
- Ceres seq_id 1565924
- Location of start within SEQ ID NO 125: at 510 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 127: from 1 to 131 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 128
- Ceres seq_id 1565925
- Location of start within SEQ ID NO 125: at 558 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 128: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
31032
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 129
- Ceres seq_id 1565926
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 130
- Ceres seq_id 1565927
- Location of start within SEQ ID NO 129: at 249 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 130: from 80 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 61
- gi No. 320563
- Description: transcription factor tga1 - Arabidopsis thaliana >gi|16550|emb|CAA48189| (X68053) transcription factor [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 368
- Location of Alignment in SEQ ID NO 130: from 1 to 368

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 131
- Ceres seq_id 1565928
- Location of start within SEQ ID NO 129: at 321 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 131: from 56 to 88 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 62
- gi No. 320563
- Description: transcription factor tga1 - Arabidopsis thaliana >gi|16550|emb|CAA48189| (X68053) transcription factor [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 368
- Location of Alignment in SEQ ID NO 131: from 1 to 344

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 132
- Ceres seq_id 1565929
- Location of start within SEQ ID NO 129: at 363 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 132: from 42 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 63
- gi No. 320563
- Description: transcription factor tga1 - Arabidopsis thaliana >gi|16550|emb|CAA48189| (X68053) transcription factor [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 368
- Location of Alignment in SEQ ID NO 132: from 1 to 330

Maximum Length Sequence:
related to:
Clone IDs:
26817
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 133
- Ceres seq_id 1565938

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 134
- Ceres seq_id 1565939
- Location of start within SEQ ID NO 133: at 599 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 134: from 1 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 64
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 90.2
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 134: from 1 to 252

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 135
- Ceres seq_id 1565940
- Location of start within SEQ ID NO 133: at 620 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 135: from 1 to 205 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 65
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 90.2
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 135: from 1 to 245

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 136
- Ceres seq_id 1565941
- Location of start within SEQ ID NO 133: at 737 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 136: from 1 to 166 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 66
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 90.2
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 136: from 1 to 206

Maximum Length Sequence:
related to:
Clone IDs:
37346
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 137
- Ceres seq_id 1565946
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 138
- Ceres seq_id 1565947
- Location of start within SEQ ID NO 137: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Globin
- Location within SEQ ID NO 138: from 4 to 154 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 67
- gi No. 2581785
- Description: (U94999) class 2 non-symbiotic hemoglobin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 138: from 1 to 158

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 139
- Ceres seq_id 1565948
- Location of start within SEQ ID NO 137: at 312 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Globin
- Location within SEQ ID NO 139: from 1 to 82 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 68
- gi No. 2581785
- Description: (U94999) class 2 non-symbiotic hemoglobin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 139: from 1 to 86

Maximum Length Sequence:
related to:
Clone IDs:
38311
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 140
- Ceres seq_id 1565957
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 141
- Ceres seq_id 1565958
- Location of start within SEQ ID NO 140: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain

- Location within SEQ ID NO 141: from 90 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 69
- gi No. 3868859
- Description: (AB013887) RAV2 [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 371
- Location of Alignment in SEQ ID NO 141: from 22 to 382

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 142
- Ceres seq_id 1565959
- Location of start within SEQ ID NO 140: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 142: from 69 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 70
- gi No. 3868859
- Description: (AB013887) RAV2 [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 371
- Location of Alignment in SEQ ID NO 142: from 1 to 361

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 143
- Ceres seq_id 1565960
- Location of start within SEQ ID NO 140: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 143: from 30 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 71
- gi No. 3868859
- Description: (AB013887) RAV2 [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 371
- Location of Alignment in SEQ ID NO 143: from 1 to 322

Maximum Length Sequence:
related to:
Clone IDs:
40627
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 144
- Ceres seq_id 1565965

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 145
- Ceres seq_id 1565966
- Location of start within SEQ ID NO 144: at 813 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)

- Location within SEQ ID NO 145: from 1 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 72
- gi No. 1707012
- Description: (U78721) tyrosyl-tRNA synthetase isolog [Arabidopsis thaliana]
- % Identity: 88.6
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 145: from 1 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 146
- Ceres seq_id 1565967
- Location of start within SEQ ID NO 144: at 846 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 146: from 1 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 73
- gi No. 1707012
- Description: (U78721) tyrosyl-tRNA synthetase isolog [Arabidopsis thaliana]
- % Identity: 88.6
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 146: from 1 to 105

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 147
- Ceres seq_id 1565968
- Location of start within SEQ ID NO 144: at 879 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 147: from 1 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 74
- gi No. 1707012
- Description: (U78721) tyrosyl-tRNA synthetase isolog [Arabidopsis thaliana]
- % Identity: 88.6
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 147: from 1 to 94

Maximum Length Sequence:
related to:
Clone IDs:
39832
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 148
- Ceres seq_id 1565973

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 149
- Ceres seq_id 1565974
- Location of start within SEQ ID NO 148: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Imidazoleglycerol-phosphate dehydratase
- Location within SEQ ID NO 149: from 121 to 265 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 75
- gi No. 462273
- Description: IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD) >gi|437213 (U02689) imidazoleglycerolphosphate dehydratase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 270
- Location of Alignment in SEQ ID NO 149: from 19 to 288

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 150
- Ceres seq_id 1565975
- Location of start within SEQ ID NO 148: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Imidazoleglycerol-phosphate dehydratase
- Location within SEQ ID NO 150: from 103 to 247 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 76
- gi No. 462273
- Description: IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD) >gi|437213 (U02689) imidazoleglycerolphosphate dehydratase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 270
- Location of Alignment in SEQ ID NO 150: from 1 to 270

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 151
- Ceres seq_id 1565976
- Location of start within SEQ ID NO 148: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Imidazoleglycerol-phosphate dehydratase
- Location within SEQ ID NO 151: from 69 to 213 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 77
- gi No. 462273
- Description: IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD) >gi|437213 (U02689) imidazoleglycerolphosphate dehydratase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 270
- Location of Alignment in SEQ ID NO 151: from 1 to 236

Maximum Length Sequence:
related to:
Clone IDs:
37887
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 152

- Ceres seq_id 1565981

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 153
- Ceres seq_id 1565982
- Location of start within SEQ ID NO 152: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 153: from 28 to 474 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 78
- gi No. 1346194
- Description: GLUTATHIONE REDUCTASE, CYTOSOLIC (GR) (GRASE) (OBP29) >gi|1022797 (U37697) glutathione reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 153: from 1 to 474

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 154
- Ceres seq_id 1565983
- Location of start within SEQ ID NO 152: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 154: from 24 to 470 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 79
- gi No. 1346194
- Description: GLUTATHIONE REDUCTASE, CYTOSOLIC (GR) (GRASE) (OBP29) >gi|1022797 (U37697) glutathione reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 154: from 1 to 470

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 155
- Ceres seq_id 1565984
- Location of start within SEQ ID NO 152: at 790 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 155: from 1 to 250 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 80
- gi No. 1346194
- Description: GLUTATHIONE REDUCTASE, CYTOSOLIC (GR) (GRASE) (OBP29) >gi|1022797 (U37697) glutathione reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 155: from 1 to 250

Maximum Length Sequence:
related to:
Clone IDs:

34275
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 156
- Ceres seq_id 1566011
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 157
- Ceres seq_id 1566012
- Location of start within SEQ ID NO 156: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 157: from 246 to 286 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 158
- Ceres seq_id 1566013
- Location of start within SEQ ID NO 156: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 158: from 216 to 256 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 159
- Ceres seq_id 1566014
- Location of start within SEQ ID NO 156: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 159: from 197 to 237 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
8742
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 160
- Ceres seq_id 1566024
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 161
- Ceres seq_id 1566025
- Location of start within SEQ ID NO 160: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 81
- gi No. 5714684
- Description: (AF169203) differentially expressed osmotic protein ODE1 [Capsicum annuum]
- % Identity: 83.1
- Alignment Length: 59

- Location of Alignment in SEQ ID NO 161: from 88 to 146

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 162
- Ceres seq_id 1566026
- Location of start within SEQ ID NO 160: at 290 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 82
- gi No. 5714684
- Description: (AF169203) differentially expressed osmotic protein ODE1 [Capsicum annuum]
- % Identity: 83.1
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 162: from 30 to 88

Maximum Length Sequence:
related to:
Clone IDs:
40175
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 163
- Ceres seq_id 1566031
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 164
- Ceres seq_id 1566032
- Location of start within SEQ ID NO 163: at 608 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Histidinol dehydrogenase
- Location within SEQ ID NO 164: from 1 to 290 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 83
- gi No. 123178
- Description: HISTIDINOL DEHYDROGENASE, CHLOROPLAST PRECURSOR (HDH) >gi|99844|pir||A39358 histidinol dehydrogenase (EC 1.1.1.23) precursor, chloroplast - cabbage >gi|167142 (M60466) histidinol dehydrogenase [Brassica oleracea]
- % Identity: 98.1
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 164: from 1 to 296

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 165
- Ceres seq_id 1566033
- Location of start within SEQ ID NO 163: at 767 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Histidinol dehydrogenase
- Location within SEQ ID NO 165: from 1 to 237 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 84
- gi No. 123178
- Description: HISTIDINOL DEHYDROGENASE, CHLOROPLAST PRECURSOR (HDH) >gi|99844|pir||A39358 histidinol dehydrogenase (EC 1.1.1.23) precursor,

chloroplast - cabbage >gi|167142 (M60466) histidinol dehydrogenase [Brassica oleracea]
- % Identity: 98.1
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 165: from 1 to 243

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 166
- Ceres seq_id 1566034
- Location of start within SEQ ID NO 163: at 845 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Histidinol dehydrogenase
- Location within SEQ ID NO 166: from 1 to 211 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 85
- gi No. 123178
- Description: HISTIDINOL DEHYDROGENASE, CHLOROPLAST PRECURSOR (HDH) >gi|99844|pir||A39358 histidinol dehydrogenase (EC 1.1.1.23) precursor, chloroplast - cabbage >gi|167142 (M60466) histidinol dehydrogenase [Brassica oleracea]
- % Identity: 98.1
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 166: from 1 to 217

Maximum Length Sequence:
related to:
Clone IDs:
31665
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 167
- Ceres seq_id 1566045
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 168
- Ceres seq_id 1566046
- Location of start within SEQ ID NO 167: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 169
- Ceres seq_id 1566047
- Location of start within SEQ ID NO 167: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 86
- gi No. 2388562
- Description: (AC000098) EST gb|ATTS0295 comes from this gene. [Arabidopsis thaliana]
- % Identity: 70.8
- Alignment Length: 73
- Location of Alignment in SEQ ID NO 169: from 1 to 71

Maximum Length Sequence:
related to:

Clone IDs:
22008
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 170
- Ceres seq_id 1566048
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 171
- Ceres seq_id 1566049
- Location of start within SEQ ID NO 170: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 171: from 12 to 268 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 172
- Ceres seq_id 1566050
- Location of start within SEQ ID NO 170: at 317 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 172: from 1 to 163 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11971
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 173
- Ceres seq_id 1566062
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 174
- Ceres seq_id 1566063
- Location of start within SEQ ID NO 173: at 42 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 174: from 12 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 87
- gi No. 2347192
- Description: (AC002338) spliceosomal protein U2B isolog [Arabidopsis thaliana]
- % Identity: 84.5
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 174: from 1 to 110

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 175
- Ceres seq_id 1566064
- Location of start within SEQ ID NO 173: at 752 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 176
- Ceres seq_id 1566065
- Location of start within SEQ ID NO 173: at 761 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
15667
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 177
- Ceres seq_id 1566100
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 178
- Ceres seq_id 1566101
- Location of start within SEQ ID NO 177: at 350 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 178: from 1 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 88
- gi No. 2129606
- Description: GTP-binding protein 3 - Arabidopsis thaliana >gi|2129701|pir||S71586 Rab11 homolog GTP-binding protein ATGB3 - Arabidopsis thaliana >gi|1184985 (U46926) ATGB3 [Arabidopsis thaliana]
- % Identity: 80.3
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 178: from 1 to 95

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 179
- Ceres seq_id 1566102
- Location of start within SEQ ID NO 177: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 179: from 1 to 90 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 89
- gi No. 2129606
- Description: GTP-binding protein 3 - Arabidopsis thaliana >gi|2129701|pir||S71586 Rab11 homolog GTP-binding protein ATGB3 - Arabidopsis thaliana >gi|1184985 (U46926) ATGB3 [Arabidopsis thaliana]
- % Identity: 80.3
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 179: from 1 to 90

Maximum Length Sequence:

related to:
Clone IDs:
92166
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 180
- Ceres seq_id 1566103
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 181
- Ceres seq_id 1566104
- Location of start within SEQ ID NO 180: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 90
- gi No. 5734718
- Description: (AC008075) F24J5.16 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 181: from 5 to 150

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 182
- Ceres seq_id 1566105
- Location of start within SEQ ID NO 180: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 91
- gi No. 5734718
- Description: (AC008075) F24J5.16 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 182: from 1 to 146

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 183
- Ceres seq_id 1566106
- Location of start within SEQ ID NO 180: at 292 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 92
- gi No. 5734718
- Description: (AC008075) F24J5.16 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 183: from 1 to 88

Maximum Length Sequence:
related to:
Clone IDs:
29099
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 184
- Ceres seq_id 1566115
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 185
- Ceres seq_id 1566116

- Location of start within SEQ ID NO 184: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/manganese superoxide dismutases (SODM)
- Location within SEQ ID NO 185: from 29 to 184 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 93
- gi No. 1351082
- Description: SUPEROXIDE DISMUTASE, CHLOROPLAST PRECURSOR (FE) >gi|166700 (M55910) Fe-superoxide dismutase [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 185: from 8 to 184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 186
- Ceres seq_id 1566117
- Location of start within SEQ ID NO 184: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/manganese superoxide dismutases (SODM)
- Location within SEQ ID NO 186: from 10 to 165 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 94
- gi No. 1351082
- Description: SUPEROXIDE DISMUTASE, CHLOROPLAST PRECURSOR (FE) >gi|166700 (M55910) Fe-superoxide dismutase [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 186: from 1 to 165

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 187
- Ceres seq_id 1566118
- Location of start within SEQ ID NO 184: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/manganese superoxide dismutases (SODM)
- Location within SEQ ID NO 187: from 1 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 95
- gi No. 1351082
- Description: SUPEROXIDE DISMUTASE, CHLOROPLAST PRECURSOR (FE) >gi|166700 (M55910) Fe-superoxide dismutase [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 187: from 1 to 139

Maximum Length Sequence:
related to:
Clone IDs:
33763
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 188
- Ceres seq_id 1566127

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 189
- Ceres seq_id 1566128
- Location of start within SEQ ID NO 188: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 189: from 6 to 52 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 96
- gi No. 2129648
- Description: MYB-related protein 33,3K - Arabidopsis thaliana >gi|1263095|emb|CAA90809| (Z54136) MYB-related protein [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 189: from 1 to 305

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 190
- Ceres seq_id 1566129
- Location of start within SEQ ID NO 188: at 508 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 97
- gi No. 2129648
- Description: MYB-related protein 33,3K - Arabidopsis thaliana >gi|1263095|emb|CAA90809| (Z54136) MYB-related protein [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 190: from 1 to 165

Maximum Length Sequence:
related to:
Clone IDs:
34897

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 191
- Ceres seq_id 1566130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 192
- Ceres seq_id 1566131
- Location of start within SEQ ID NO 191: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L27 protein
- Location within SEQ ID NO 192: from 47 to 125 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 98
- gi No. 1173010
- Description: CHLOROPLAST 50S RIBOSOMAL PROTEIN L27 >gi|999433|dbj|BAA05097| (D26101) ribosomal protein L27 [Porphyridium cruentum]
- % Identity: 70.2
- Alignment Length: 47
- Location of Alignment in SEQ ID NO 192: from 54 to 100

Maximum Length Sequence:
related to:
Clone IDs:
33917
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 193
- Ceres seq_id 1566132
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 194
- Ceres seq_id 1566133
- Location of start within SEQ ID NO 193: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lactate/malate dehydrogenase
- Location within SEQ ID NO 194: from 76 to 381 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 99
- gi No. 3929649
- Description: (AJ131205) mitochondrial NAD-dependent malate dehydrogenase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 341
- Location of Alignment in SEQ ID NO 194: from 46 to 386

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 195
- Ceres seq_id 1566134
- Location of start within SEQ ID NO 193: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lactate/malate dehydrogenase
- Location within SEQ ID NO 195: from 31 to 336 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 100
- gi No. 3929649
- Description: (AJ131205) mitochondrial NAD-dependent malate dehydrogenase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 341
- Location of Alignment in SEQ ID NO 195: from 1 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 196
- Ceres seq_id 1566135
- Location of start within SEQ ID NO 193: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lactate/malate dehydrogenase
- Location within SEQ ID NO 196: from 27 to 332 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 101
- gi No. 3929649
- Description: (AJ131205) mitochondrial NAD-dependent malate dehydrogenase [Arabidopsis thaliana]

- % Identity: 99.7
- Alignment Length: 341
- Location of Alignment in SEQ ID NO 196: from 1 to 337

Maximum Length Sequence:
related to:
Clone IDs:
40049
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 197
- Ceres seq_id 1566150
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 198
- Ceres seq_id 1566151
- Location of start within SEQ ID NO 197: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 102
- gi No. 2829923
- Description: (AC002291) Similar to uridylyl transferases [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 453
- Location of Alignment in SEQ ID NO 198: from 1 to 453

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 199
- Ceres seq_id 1566152
- Location of start within SEQ ID NO 197: at 385 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 103
- gi No. 2829923
- Description: (AC002291) Similar to uridylyl transferases [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 453
- Location of Alignment in SEQ ID NO 199: from 1 to 411

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 200
- Ceres seq_id 1566153
- Location of start within SEQ ID NO 197: at 484 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 104
- gi No. 2829923
- Description: (AC002291) Similar to uridylyl transferases [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 453
- Location of Alignment in SEQ ID NO 200: from 1 to 378

Maximum Length Sequence:
related to:

Clone IDs:
32703
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 201
- Ceres seq_id 1566154
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 202
- Ceres seq_id 1566155
- Location of start within SEQ ID NO 201: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 202: from 8 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 105
- gi No. 544425
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 85.2
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 202: from 1 to 145

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 203
- Ceres seq_id 1566156
- Location of start within SEQ ID NO 201: at 225 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 203: from 30 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 106
- gi No. 544425
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 85.2
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 203: from 1 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 204
- Ceres seq_id 1566157
- Location of start within SEQ ID NO 201: at 246 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 204: from 23 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 107
- gi No. 544425

- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 85.2
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 204: from 1 to 79

Maximum Length Sequence:
related to:
Clone IDs:
42826
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 205
- Ceres seq_id 1566170
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 206
- Ceres seq_id 1566171
- Location of start within SEQ ID NO 205: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectate lyase
- Location within SEQ ID NO 206: from 138 to 402 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 108
- gi No. 280407
- Description: allergen Amb a I - common ragweed (fragments)
- % Identity: 71.9
- Alignment Length: 32
- Location of Alignment in SEQ ID NO 206: from 288 to 319

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 207
- Ceres seq_id 1566173
- Location of start within SEQ ID NO 205: at 55 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectate lyase
- Location within SEQ ID NO 207: from 120 to 384 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 109
- gi No. 280407
- Description: allergen Amb a I - common ragweed (fragments)
- % Identity: 71.9
- Alignment Length: 32
- Location of Alignment in SEQ ID NO 207: from 270 to 301

Maximum Length Sequence:
related to:
Clone IDs:
40893
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 208
- Ceres seq_id 1566181
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 209
- Ceres seq_id 1566182

- Location of start within SEQ ID NO 208: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphopantetheine attachment site
- Location within SEQ ID NO 209: from 107 to 174 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 110
- gi No. 3776558
- Description: (AC005388) Identical to gb|L14814 DNA for tissue-specific acyl carrier protein isoform 2 from A. thaliana. ESTs gb|AA597351, gb|T41805, gb|H36871, gb|R30210, gb|AA042549, gb|Z47650, gb|H76304
- % Identity: 99.3
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 209: from 48 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 210
- Ceres seq_id 1566183
- Location of start within SEQ ID NO 208: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 211
- Ceres seq_id 1566184
- Location of start within SEQ ID NO 208: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphopantetheine attachment site
- Location within SEQ ID NO 211: from 60 to 127 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 111
- gi No. 3776558
- Description: (AC005388) Identical to gb|L14814 DNA for tissue-specific acyl carrier protein isoform 2 from A. thaliana. ESTs gb|AA597351, gb|T41805, gb|H36871, gb|R30210, gb|AA042549, gb|Z47650, gb|H76304
- % Identity: 99.3
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 211: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
39258
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 212
- Ceres seq_id 1566185
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 213
- Ceres seq_id 1566186
- Location of start within SEQ ID NO 212: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 112
- gi No. 3377825
- Description: (AF075598) contains similarity to fibrillins [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 213: from 6 to 322

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 214
- Ceres seq_id 1566187
- Location of start within SEQ ID NO 212: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 113
- gi No. 3377825
- Description: (AF075598) contains similarity to fibrillins [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 214: from 1 to 317

Maximum Length Sequence:
related to:
Clone IDs:
10375
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 215
- Ceres seq_id 1566214
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 216
- Ceres seq_id 1566215
- Location of start within SEQ ID NO 215: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 114
- gi No. 2462081
- Description: (Y09427) squamosa-promoter binding protein like 3 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 131
- Location of Alignment in SEQ ID NO 216: from 1 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 217
- Ceres seq_id 1566216
- Location of start within SEQ ID NO 215: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 115
- gi No. 2462081
- Description: (Y09427) squamosa-promoter binding protein like 3 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 131

- Location of Alignment in SEQ ID NO 217: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
33767
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 218
- Ceres seq_id 1566220
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 219
- Ceres seq_id 1566221
- Location of start within SEQ ID NO 218: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 219: from 40 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 116
- gi No. 119111
- Description: EBNA-2 NUCLEAR PROTEIN >gi|1083964|pir||S42442 EBNA2 protein - human herpesvirus 4 >gi|1632787|emb|CAA24877.1| (V01555) BYRF1, encodes EBNA-2 (Dambaugh et al, 1984; Dillner et al, 1984) [Human herpesvirus 4]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 219: from 106 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 220
- Ceres seq_id 1566222
- Location of start within SEQ ID NO 218: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 220: from 15 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 117
- gi No. 119111
- Description: EBNA-2 NUCLEAR PROTEIN >gi|1083964|pir||S42442 EBNA2 protein - human herpesvirus 4 >gi|1632787|emb|CAA24877.1| (V01555) BYRF1, encodes EBNA-2 (Dambaugh et al, 1984; Dillner et al, 1984) [Human herpesvirus 4]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 220: from 81 to 91

Maximum Length Sequence:
related to:
Clone IDs:
34126
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 221
- Ceres seq_id 1566223
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 222
- Ceres seq_id 1566224

- Location of start within SEQ ID NO 221: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 118
- gi No. 461453
- Description: AUXIN-BINDING PROTEIN 1 PRECURSOR (ABP)
>gi|322467|pir||S31584 auxin-binding protein precursor - Arabidopsis thaliana
>gi|16201|emb|CAA49526| (X69901) auxin-binding protein [Arabidopsis thaliana]
>gi|251899|bbs|109280 (S40550)
- % Identity: 99.5
- Alignment Length: 198
- Location of Alignment in SEQ ID NO 222: from 11 to 208

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 223
- Ceres seq_id 1566225
- Location of start within SEQ ID NO 221: at 31 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 119
- gi No. 461453
- Description: AUXIN-BINDING PROTEIN 1 PRECURSOR (ABP)
>gi|322467|pir||S31584 auxin-binding protein precursor - Arabidopsis thaliana
>gi|16201|emb|CAA49526| (X69901) auxin-binding protein [Arabidopsis thaliana]
>gi|251899|bbs|109280 (S40550)
- % Identity: 99.5
- Alignment Length: 198
- Location of Alignment in SEQ ID NO 223: from 1 to 198

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 224
- Ceres seq_id 1566226
- Location of start within SEQ ID NO 221: at 217 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 120
- gi No. 461453
- Description: AUXIN-BINDING PROTEIN 1 PRECURSOR (ABP)
>gi|322467|pir||S31584 auxin-binding protein precursor - Arabidopsis thaliana
>gi|16201|emb|CAA49526| (X69901) auxin-binding protein [Arabidopsis thaliana]
>gi|251899|bbs|109280 (S40550)
- % Identity: 99.5
- Alignment Length: 198
- Location of Alignment in SEQ ID NO 224: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
42965
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 225
- Ceres seq_id 1566234
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 226
- Ceres seq_id 1566235

- Location of start within SEQ ID NO 225: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glucosamine-6-phosphate isomerase
- Location within SEQ ID NO 226: from 111 to 319 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 227
- Ceres seq_id 1566236
- Location of start within SEQ ID NO 225: at 318 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glucosamine-6-phosphate isomerase
- Location within SEQ ID NO 227: from 42 to 250 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
42925
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 228
- Ceres seq_id 1566237
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 229
- Ceres seq_id 1566238
- Location of start within SEQ ID NO 228: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 121
- gi No. 5733877
- Description: (AC007932) F11A17.13 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 229: from 33 to 156

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 230
- Ceres seq_id 1566239
- Location of start within SEQ ID NO 228: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 122
- gi No. 5733877
- Description: (AC007932) F11A17.13 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 230: from 1 to 124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 231
- Ceres seq_id 1566240

- Location of start within SEQ ID NO 228: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 123
- gi No. 5733877
- Description: (AC007932) F11A17.13 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 231: from 1 to 108

Maximum Length Sequence:
related to:
Clone IDs:
33591

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 232
- Ceres seq_id 1566241

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 233
- Ceres seq_id 1566242
- Location of start within SEQ ID NO 232: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 233: from 59 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 124
- gi No. 3914425
- Description: PROTEASOME EPSILON CHAIN PRECURSOR (MACROPAIN EPSILON CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX EPSILON CHAIN) >gi|2511596|emb|CAA74029.1| (Y13695) multicatalytic endopeptidase >gi|3421117 (AF043536) 20S proteasome beta subunit PBE1
- % Identity: 100
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 233: from 1 to 212

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 234
- Ceres seq_id 1566243
- Location of start within SEQ ID NO 232: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 234: from 48 to 190 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 125
- gi No. 3914425
- Description: PROTEASOME EPSILON CHAIN PRECURSOR (MACROPAIN EPSILON CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX EPSILON CHAIN) >gi|2511596|emb|CAA74029.1| (Y13695) multicatalytic endopeptidase >gi|3421117 (AF043536) 20S proteasome beta subunit PBE1
- % Identity: 100
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 234: from 1 to 201

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 235
- Ceres seq_id 1566244
- Location of start within SEQ ID NO 232: at 152 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 235: from 46 to 188 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 126
- gi No. 3914425
- Description: PROTEASOME EPSILON CHAIN PRECURSOR (MACROPAIN EPSILON CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX EPSILON CHAIN) >gi|2511596|emb|CAA74029.1| (Y13695) multicatalytic endopeptidase >gi|3421117 (AF043536) 20S proteasome beta subunit PBE1
- % Identity: 100
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 235: from 1 to 199

Maximum Length Sequence:
related to:
Clone IDs:
15914
1042
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 236
- Ceres seq_id 1566253
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 237
- Ceres seq_id 1566254
- Location of start within SEQ ID NO 236: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DEAD/DEAH box helicase
- Location within SEQ ID NO 237: from 44 to 256 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 238
- Ceres seq_id 1566255
- Location of start within SEQ ID NO 236: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DEAD/DEAH box helicase
- Location within SEQ ID NO 238: from 2 to 214 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 239
- Ceres seq_id 1566256
- Location of start within SEQ ID NO 236: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DEAD/DEAH box helicase

- Location within SEQ ID NO 239: from 1 to 207 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
92869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 240
- Ceres seq_id 1566257
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 241
- Ceres seq_id 1566258
- Location of start within SEQ ID NO 240: at 181 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 241: from 7 to 72 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 242
- Ceres seq_id 1566259
- Location of start within SEQ ID NO 240: at 322 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
93174
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 243
- Ceres seq_id 1566268
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 244
- Ceres seq_id 1566269
- Location of start within SEQ ID NO 243: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S21e
- Location within SEQ ID NO 244: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 127
- gi No. 2500497
- Description: 40S RIBOSOMAL PROTEIN S21
>gi|1419372|emb|CAA67225.1| (X98656) ribosomal protein S21 [Zea mays]
- % Identity: 81.5
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 244: from 1 to 81

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 245
- Ceres seq_id 1566270

- Location of start within SEQ ID NO 243: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 246
- Ceres seq_id 1566271
- Location of start within SEQ ID NO 243: at 206 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
93427
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 247
- Ceres seq_id 1566276
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 248
- Ceres seq_id 1566277
- Location of start within SEQ ID NO 247: at 43 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin light chain
- Location within SEQ ID NO 248: from 5 to 205 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 249
- Ceres seq_id 1566278
- Location of start within SEQ ID NO 247: at 316 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin light chain
- Location within SEQ ID NO 249: from 1 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 250
- Ceres seq_id 1566279
- Location of start within SEQ ID NO 247: at 403 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin light chain
- Location within SEQ ID NO 250: from 1 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

93996
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 251
- Ceres seq_id 1566290
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 252
- Ceres seq_id 1566291
- Location of start within SEQ ID NO 251: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 252: from 289 to 361 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 253
- Ceres seq_id 1566292
- Location of start within SEQ ID NO 251: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 253: from 272 to 344 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 254
- Ceres seq_id 1566293
- Location of start within SEQ ID NO 251: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 254: from 270 to 342 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
94231
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 255
- Ceres seq_id 1566297
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 256
- Ceres seq_id 1566298
- Location of start within SEQ ID NO 255: at 54 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 128
- gi No. 2618727
- Description: (U49075) IAA19 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 256: from 53 to 169

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 257
- Ceres seq_id 1566299
- Location of start within SEQ ID NO 255: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 129
- gi No. 2618727
- Description: (U49075) IAA19 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 257: from 28 to 144

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 258
- Ceres seq_id 1566300
- Location of start within SEQ ID NO 255: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 130
- gi No. 2618727
- Description: (U49075) IAA19 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 258: from 27 to 143

Maximum Length Sequence:
related to:
Clone IDs:
18482
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 259
- Ceres seq_id 1566353

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 260
- Ceres seq_id 1566354
- Location of start within SEQ ID NO 259: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 131
- gi No. 5678827
- Description: (AF151646) PC-MYB1 [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 484
- Location of Alignment in SEQ ID NO 260: from 1 to 484

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 261
- Ceres seq_id 1566355
- Location of start within SEQ ID NO 259: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 132
- gi No. 5678827
- Description: (AF151646) PC-MYB1 [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 484
- Location of Alignment in SEQ ID NO 261: from 1 to 412

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 262
- Ceres seq_id 1566356
- Location of start within SEQ ID NO 259: at 375 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 133
- gi No. 5678827
- Description: (AF151646) PC-MYB1 [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 484
- Location of Alignment in SEQ ID NO 262: from 1 to 360

Maximum Length Sequence:
related to:
Clone IDs:
20633

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 263
- Ceres seq_id 1566364

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 264
- Ceres seq_id 1566365
- Location of start within SEQ ID NO 263: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 264: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 134
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 264: from 1 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 265
- Ceres seq_id 1566366
- Location of start within SEQ ID NO 263: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 265: from 1 to 68 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 135
- gi No. 2252857

- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 265: from 1 to 68

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 266
- Ceres seq_id 1566367
- Location of start within SEQ ID NO 263: at 216 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 266: from 1 to 63 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 136
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 266: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
18804
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 267
- Ceres seq_id 1566379

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 268
- Ceres seq_id 1566380
- Location of start within SEQ ID NO 267: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 137
- gi No. 2160154
- Description: (AC000375) F19K23.18 gene product [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 431
- Location of Alignment in SEQ ID NO 268: from 1 to 431

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 269
- Ceres seq_id 1566381
- Location of start within SEQ ID NO 267: at 8 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 138
- gi No. 2160154
- Description: (AC000375) F19K23.18 gene product [Arabidopsis thaliana]
- % Identity: 99.1

- Alignment Length: 431
- Location of Alignment in SEQ ID NO 269: from 1 to 429

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 270
- Ceres seq_id 1566382
- Location of start within SEQ ID NO 267: at 41 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 139
- gi No. 2160154
- Description: (AC000375) F19K23.18 gene product [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 431
- Location of Alignment in SEQ ID NO 270: from 1 to 418

Maximum Length Sequence:
related to:
Clone IDs:
22919
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 271
- Ceres seq_id 1566387
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 272
- Ceres seq_id 1566388
- Location of start within SEQ ID NO 271: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ribosomal L5P family C-terminus
- Location within SEQ ID NO 272: from 157 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 140
- gi No. 2191128
- Description: (AF007269) belongs to the L5P family of ribosomal proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 272: from 19 to 280

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 273
- Ceres seq_id 1566389
- Location of start within SEQ ID NO 271: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ribosomal L5P family C-terminus
- Location within SEQ ID NO 273: from 139 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 141
- gi No. 2191128
- Description: (AF007269) belongs to the L5P family of ribosomal proteins [Arabidopsis thaliana]
- % Identity: 100

- Alignment Length: 262
- Location of Alignment in SEQ ID NO 273: from 1 to 262

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 274
- Ceres seq_id 1566390
- Location of start within SEQ ID NO 271: at 141 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ribosomal L5P family C-terminus
- Location within SEQ ID NO 274: from 111 to 205 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 142
- gi No. 2191128
- Description: (AF007269) belongs to the L5P family of ribosomal proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 274: from 1 to 234

Maximum Length Sequence:
related to:
Clone IDs:
25723
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 275
- Ceres seq_id 1566395

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 276
- Ceres seq_id 1566396
- Location of start within SEQ ID NO 275: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 276: from 33 to 133 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 143
- gi No. 3329623
- Description: (AF078790) No definition line found [Caenorhabditis elegans]
- % Identity: 90.3
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 276: from 106 to 136

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 277
- Ceres seq_id 1566397
- Location of start within SEQ ID NO 275: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 277: from 1 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 144
- gi No. 3329623

- Description: (AF078790) No definition line found [Caenorhabditis elegans]
- % Identity: 90.3
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 277: from 73 to 103

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 278
- Ceres seq_id 1566398
- Location of start within SEQ ID NO 275: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 278: from 1 to 89 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 145
- gi No. 3329623
- Description: (AF078790) No definition line found [Caenorhabditis elegans]
- % Identity: 90.3
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 278: from 62 to 92

Maximum Length Sequence:
related to:
Clone IDs:
32558
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 279
- Ceres seq_id 1566423

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 280
- Ceres seq_id 1566424
- Location of start within SEQ ID NO 279: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 280: from 101 to 212 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 281
- Ceres seq_id 1566425
- Location of start within SEQ ID NO 279: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 281: from 78 to 189 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 282
- Ceres seq_id 1566426
- Location of start within SEQ ID NO 279: at 199 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 282: from 58 to 169 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
26427
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 283
- Ceres seq_id 1566443
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 284
- Ceres seq_id 1566444
- Location of start within SEQ ID NO 283: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 284: from 64 to 174 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 146
- gi No. 123541
- Description: 17.4 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.4)
>gi|81639|pir||JQ0351 heat shock protein 17 - Arabidopsis thaliana
>gi|16340|emb|CAA35182| (X17293) heat shock protein (AA 1 - 156) [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 156
- Location of Alignment in SEQ ID NO 284: from 20 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 285
- Ceres seq_id 1566445
- Location of start within SEQ ID NO 283: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 285: from 45 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 147
- gi No. 123541
- Description: 17.4 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.4)
>gi|81639|pir||JQ0351 heat shock protein 17 - Arabidopsis thaliana
>gi|16340|emb|CAA35182| (X17293) heat shock protein (AA 1 - 156) [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 156
- Location of Alignment in SEQ ID NO 285: from 1 to 156

Maximum Length Sequence:
related to:
Clone IDs:
36364
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 286
- Ceres seq_id 1566453

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 287
- Ceres seq_id 1566454
- Location of start within SEQ ID NO 286: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 287: from 135 to 221 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 288
- Ceres seq_id 1566455
- Location of start within SEQ ID NO 286: at 171 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 288: from 120 to 206 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 289
- Ceres seq_id 1566456
- Location of start within SEQ ID NO 286: at 291 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 289: from 80 to 166 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37870

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 290
- Ceres seq_id 1566468

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 291
- Ceres seq_id 1566469
- Location of start within SEQ ID NO 290: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lyase
- Location within SEQ ID NO 291: from 95 to 515 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 148
- gi No. 1769568
- Description: (U82202) fumarase; fumarate hydratase [Arabidopsis thaliana]
- % Identity: 98.8

- Alignment Length: 423
- Location of Alignment in SEQ ID NO 291: from 100 to 522

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 292
- Ceres seq_id 1566470
- Location of start within SEQ ID NO 290: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lyase
- Location within SEQ ID NO 292: from 66 to 486 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 149
- gi No. 1769568
- Description: (U82202) fumarase; fumarate hydratase [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 423
- Location of Alignment in SEQ ID NO 292: from 71 to 493

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 293
- Ceres seq_id 1566471
- Location of start within SEQ ID NO 290: at 300 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lyase
- Location within SEQ ID NO 293: from 1 to 416 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 150
- gi No. 1769568
- Description: (U82202) fumarase; fumarate hydratase [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 423
- Location of Alignment in SEQ ID NO 293: from 1 to 423

Maximum Length Sequence:
related to:
Clone IDs:
38663
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 294
- Ceres seq_id 1566505

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 295
- Ceres seq_id 1566506
- Location of start within SEQ ID NO 294: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 295: from 23 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 151
- gi No. 70644

- Description: ubiquitin precursor - common sunflower (fragment)
- % Identity: 94.4
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 295: from 23 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 296
- Ceres seq_id 1566507
- Location of start within SEQ ID NO 294: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 296: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 152
- gi No. 70644
- Description: ubiquitin precursor - common sunflower (fragment)
- % Identity: 94.4
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 296: from 1 to 133

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 297
- Ceres seq_id 1566508
- Location of start within SEQ ID NO 294: at 527 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 297: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 153
- gi No. 70644
- Description: ubiquitin precursor - common sunflower (fragment)
- % Identity: 94.4
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 297: from 1 to 153

Maximum Length Sequence:
related to:
Clone IDs:
36303
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 298
- Ceres seq_id 1566517

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 299
- Ceres seq_id 1566518
- Location of start within SEQ ID NO 298: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 299: from 116 to 187 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 154
- gi No. 1076303

- Description: RNA-binding protein cp29 precursor - Arabidopsis thaliana >gi|681902|dbj|BAA06518| (D31710) cp29 [Arabidopsis thaliana]
- % Identity: 72.9
- Alignment Length: 210
- Location of Alignment in SEQ ID NO 299: from 28 to 225

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 300
- Ceres seq_id 1566519
- Location of start within SEQ ID NO 298: at 425 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 300: from 88 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 155
- gi No. 1076303
- Description: RNA-binding protein cp29 precursor - Arabidopsis thaliana >gi|681902|dbj|BAA06518| (D31710) cp29 [Arabidopsis thaliana]
- % Identity: 72.9
- Alignment Length: 210
- Location of Alignment in SEQ ID NO 300: from 1 to 84

Maximum Length Sequence:
related to:
Clone IDs:
38630
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 301
- Ceres seq_id 1566531

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 302
- Ceres seq_id 1566532
- Location of start within SEQ ID NO 301: at 232 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 302: from 1 to 91 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 303
- Ceres seq_id 1566533
- Location of start within SEQ ID NO 301: at 271 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 303: from 1 to 78 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 304
- Ceres seq_id 1566534
- Location of start within SEQ ID NO 301: at 295 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 304: from 1 to 70 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6982
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 305
- Ceres seq_id 1566553
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 306
- Ceres seq_id 1566554
- Location of start within SEQ ID NO 305: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 306: from 12 to 40 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 156
- gi No. 1168748
- Description: CALMODULIN-4 >gi|479693|pir||S35185 calmodulin 4 - Arabidopsis thaliana >gi|16223|emb|CAA78057| (Z12022) calmodulin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 306: from 1 to 149

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 307
- Ceres seq_id 1566555
- Location of start within SEQ ID NO 305: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 157
- gi No. 1168748
- Description: CALMODULIN-4 >gi|479693|pir||S35185 calmodulin 4 - Arabidopsis thaliana >gi|16223|emb|CAA78057| (Z12022) calmodulin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 307: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 308
- Ceres seq_id 1566556
- Location of start within SEQ ID NO 305: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 158
- gi No. 1168748

- Description: CALMODULIN-4 >gi|479693|pir||S35185 calmodulin 4 - Arabidopsis thaliana >gi|16223|emb|CAA78057| (Z12022) calmodulin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 149
- Location of Alignment in SEQ ID NO 308: from 1 to 98

Maximum Length Sequence:
related to:
Clone IDs:
19785
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 309
- Ceres seq_id 1566580
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 310
- Ceres seq_id 1566581
- Location of start within SEQ ID NO 309: at 581 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 310: from 125 to 163 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 159
- gi No. 3193324
- Description: (AF069299) contains similarity to WD domains, G-beta repeats (Pfam: G-beta.hmm, score: 22.80 and 35.84) [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 310: from 1 to 260

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 311
- Ceres seq_id 1566582
- Location of start within SEQ ID NO 309: at 602 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 311: from 118 to 156 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 160
- gi No. 3193324
- Description: (AF069299) contains similarity to WD domains, G-beta repeats (Pfam: G-beta.hmm, score: 22.80 and 35.84) [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 311: from 1 to 253

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 312
- Ceres seq_id 1566583
- Location of start within SEQ ID NO 309: at 848 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 312: from 36 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 161
- gi No. 3193324
- Description: (AF069299) contains similarity to WD domains, G-beta repeats (Pfam: G-beta.hmm, score: 22.80 and 35.84) [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 312: from 1 to 171

Maximum Length Sequence:
related to:
Clone IDs:
19319
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 313
- Ceres seq_id 1566588
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 314
- Ceres seq_id 1566589
- Location of start within SEQ ID NO 313: at 26 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oleosin
- Location within SEQ ID NO 314: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 162
- gi No. 2129657
- Description: oleosin isoform - Arabidopsis thaliana >gi|987014|emb|CAA90877| (Z54164) oleosin [Arabidopsis thaliana] >gi|987016|emb|CAA90878| (Z54165) oleosin [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 80
- Location of Alignment in SEQ ID NO 314: from 1 to 80

Maximum Length Sequence:
related to:
Clone IDs:
95453
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 315
- Ceres seq_id 1566597
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 316
- Ceres seq_id 1566598
- Location of start within SEQ ID NO 315: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 316: from 10 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 163
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]

- % Identity: 95.5
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 316: from 1 to 169

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 317
- Ceres seq_id 1566599
- Location of start within SEQ ID NO 315: at 244 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 317: from 21 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 164
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 95.5
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 317: from 1 to 108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 318
- Ceres seq_id 1566600
- Location of start within SEQ ID NO 315: at 284 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
95677

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 319
- Ceres seq_id 1566608

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 320
- Ceres seq_id 1566609
- Location of start within SEQ ID NO 319: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 320: from 30 to 468 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 321
- Ceres seq_id 1566610
- Location of start within SEQ ID NO 319: at 189 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450

- Location within SEQ ID NO 321: from 1 to 406 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 322
- Ceres seq_id 1566611
- Location of start within SEQ ID NO 319: at 333 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 322: from 1 to 358 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
10879
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 323
- Ceres seq_id 1566616
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 324
- Ceres seq_id 1566617
- Location of start within SEQ ID NO 323: at 54 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 324: from 7 to 85 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 165
- gi No. 1816635
- Description: (U85614) SRG3 [Mus musculus]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 324: from 88 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 325
- Ceres seq_id 1566618
- Location of start within SEQ ID NO 323: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 325: from 5 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 166
- gi No. 1816635
- Description: (U85614) SRG3 [Mus musculus]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 325: from 86 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 326

- Ceres seq_id 1566619
- Location of start within SEQ ID NO 323: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 326: from 1 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 167
- gi No. 1816635
- Description: (U85614) SRG3 [Mus musculus]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 326: from 80 to 91

Maximum Length Sequence:
related to:
Clone IDs:
96723
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 327
- Ceres seq_id 1566630
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 328
- Ceres seq_id 1566631
- Location of start within SEQ ID NO 327: at 105 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 168
- gi No. 2462077
- Description: (Y11871) Oxa1 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 328: from 1 to 429

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 329
- Ceres seq_id 1566632
- Location of start within SEQ ID NO 327: at 351 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 169
- gi No. 2462077
- Description: (Y11871) Oxa1 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 329: from 1 to 347

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 330
- Ceres seq_id 1566633
- Location of start within SEQ ID NO 327: at 393 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 170
- gi No. 2462077
- Description: (Y11871) Oxa1 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 330: from 1 to 333

Maximum Length Sequence:
related to:
Clone IDs:
96896
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 331
- Ceres seq_id 1566641
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 332
- Ceres seq_id 1566642
- Location of start within SEQ ID NO 331: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 333
- Ceres seq_id 1566643
- Location of start within SEQ ID NO 331: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 334
- Ceres seq_id 1566644
- Location of start within SEQ ID NO 331: at 361 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AN1-like Zinc finger
- Location within SEQ ID NO 334: from 1 to 20 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1097
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 335
- Ceres seq_id 1566650
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 336
- Ceres seq_id 1566651
- Location of start within SEQ ID NO 335: at 37 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L35Ae
- Location within SEQ ID NO 336: from 11 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 171
- gi No. 1709990
- Description: 60S RIBOSOMAL PROTEIN L35A
- % Identity: 98
- Alignment Length: 50
- Location of Alignment in SEQ ID NO 336: from 1 to 50

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 337
- Ceres seq_id 1566652
- Location of start within SEQ ID NO 335: at 369 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
97415
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 338
- Ceres seq_id 1566661

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 339
- Ceres seq_id 1566662
- Location of start within SEQ ID NO 338: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 339: from 43 to 152 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 172
- gi No. 4587546
- Description: (AC006577) Belongs to the PF|00011 Hsp20/alpha crystallin family. EST gb|W4312 comes from this gene. [Arabidopsis thaliana]
- % Identity: 98.1
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 339: from 1 to 155

Maximum Length Sequence:
related to:
Clone IDs:
97434
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 340
- Ceres seq_id 1566663

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 341
- Ceres seq_id 1566664
- Location of start within SEQ ID NO 340: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 341: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 173
- gi No. 1710591
- Description: 60S ACIDIC RIBOSOMAL PROTEIN P2
- % Identity: 76.6
- Alignment Length: 77
- Location of Alignment in SEQ ID NO 341: from 39 to 113

Maximum Length Sequence:
related to:
Clone IDs:
10984
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 342
- Ceres seq_id 1566665
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 343
- Ceres seq_id 1566666
- Location of start within SEQ ID NO 342: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 174
- gi No. 929566
- Description: (X05285) fibrillarin [Drosophila melanogaster]
- % Identity: 73.5
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 343: from 7 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 344
- Ceres seq_id 1566667
- Location of start within SEQ ID NO 342: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 175
- gi No. 929566
- Description: (X05285) fibrillarin [Drosophila melanogaster]
- % Identity: 73.5
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 344: from 3 to 49

Maximum Length Sequence:
related to:
Clone IDs:
97535
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 345
- Ceres seq_id 1566671
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 346
- Ceres seq_id 1566672
- Location of start within SEQ ID NO 345: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 346: from 1 to 279 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 176
- gi No. 1495804
- Description: (X96406) 13-lipoxygenase [Solanum tuberosum]
- % Identity: 79.2
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 346: from 1 to 279

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 347
- Ceres seq_id 1566673
- Location of start within SEQ ID NO 345: at 212 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 347: from 1 to 209 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 177
- gi No. 1495804
- Description: (X96406) 13-lipoxygenase [Solanum tuberosum]
- % Identity: 79.2
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 347: from 1 to 209

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 348
- Ceres seq_id 1566674
- Location of start within SEQ ID NO 345: at 807 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 348: from 2 to 224 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 178
- gi No. 1495804
- Description: (X96406) 13-lipoxygenase [Solanum tuberosum]
- % Identity: 74.2
- Alignment Length: 240
- Location of Alignment in SEQ ID NO 348: from 2 to 240

Maximum Length Sequence:
related to:
Clone IDs:
11009
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 349
- Ceres seq_id 1566683

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 350
- Ceres seq_id 1566684
- Location of start within SEQ ID NO 349: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Purine/pyrimidine phosphoribosyl transferases
- Location within SEQ ID NO 350: from 15 to 172 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 179
- gi No. 3892712
- Description: (AL033545) adenine phosphoribosyltransferase (EC 2.4.2.7)-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 183
- Location of Alignment in SEQ ID NO 350: from 1 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 351
- Ceres seq_id 1566685
- Location of start within SEQ ID NO 349: at 244 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Purine/pyrimidine phosphoribosyl transferases
- Location within SEQ ID NO 351: from 1 to 142 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 180
- gi No. 3892712
- Description: (AL033545) adenine phosphoribosyltransferase (EC 2.4.2.7)-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 183
- Location of Alignment in SEQ ID NO 351: from 1 to 153

Maximum Length Sequence:
related to:
Clone IDs:
1105
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 352
- Ceres seq_id 1566691
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 353
- Ceres seq_id 1566692
- Location of start within SEQ ID NO 352: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 181
- gi No. 2982299
- Description: (AF051234) transcription factor BTF3 homolog [Picea mariana]
- % Identity: 75
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 353: from 63 to 186

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 354
- Ceres seq_id 1566693
- Location of start within SEQ ID NO 352: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 182
- gi No. 2982299

- Description: (AF051234) transcription factor BTF3 homolog [Picea mariana]
- % Identity: 75
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 354: from 34 to 157

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 355
- Ceres seq_id 1566694
- Location of start within SEQ ID NO 352: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 183
- gi No. 2982299
- Description: (AF051234) transcription factor BTF3 homolog [Picea mariana]
- % Identity: 75
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 355: from 28 to 151

Maximum Length Sequence:
related to:
Clone IDs:
98232
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 356
- Ceres seq_id 1566699
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 357
- Ceres seq_id 1566700
- Location of start within SEQ ID NO 356: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 184
- gi No. 2190548
- Description: (AC001229) EST gb|ATTS1121 comes from this gene. [Arabidopsis thaliana]
- % Identity: 72.9
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 357: from 33 to 91

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 358
- Ceres seq_id 1566701
- Location of start within SEQ ID NO 356: at 279 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 185
- gi No. 2190548
- Description: (AC001229) EST gb|ATTS1121 comes from this gene. [Arabidopsis thaliana]
- % Identity: 72.9
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 358: from 2 to 60

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 359
- Ceres seq_id 1566702
- Location of start within SEQ ID NO 356: at 332 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
98545
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 360
- Ceres seq_id 1566707
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 361
- Ceres seq_id 1566708
- Location of start within SEQ ID NO 360: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 186
- gi No. 4115934
- Description: (AF118223) contains similarity to Methanobacterium thermoautotrophicum transcriptional regulator (GB:AE000850) [Arabidopsis thaliana]
- % Identity: 84
- Alignment Length: 25
- Location of Alignment in SEQ ID NO 361: from 16 to 40

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 362
- Ceres seq_id 1566709
- Location of start within SEQ ID NO 360: at 697 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 187
- gi No. 4115935
- Description: (AF118223) contains similarity to Helicobacter pylori peptide methionine sulfoxide reductase (msrA) (GB:AE000542) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 51
- Location of Alignment in SEQ ID NO 362: from 1 to 45

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 363
- Ceres seq_id 1566710
- Location of start within SEQ ID NO 360: at 737 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
98672
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 364
- Ceres seq_id 1566711
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 365
- Ceres seq_id 1566712
- Location of start within SEQ ID NO 364: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 365: from 285 to 323 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 366
- Ceres seq_id 1566713
- Location of start within SEQ ID NO 364: at 203 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 366: from 263 to 301 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 367
- Ceres seq_id 1566714
- Location of start within SEQ ID NO 364: at 338 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 367: from 218 to 256 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
98677
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 368
- Ceres seq_id 1566715
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 369
- Ceres seq_id 1566716
- Location of start within SEQ ID NO 368: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 369: from 210 to 344 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 370
- Ceres seq_id 1566717
- Location of start within SEQ ID NO 368: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 370: from 187 to 321 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 371
- Ceres seq_id 1566718
- Location of start within SEQ ID NO 368: at 199 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 371: from 167 to 301 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11114
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 372
- Ceres seq_id 1566719

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 373
- Ceres seq_id 1566720
- Location of start within SEQ ID NO 372: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 188
- gi No. 2135627
- Description: metallothionein old gene name 'MT2' - human (fragment) >gi|187528 (M26637) metallothionein old gene name 'MT2' [Homo sapiens]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 373: from 54 to 71

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 374
- Ceres seq_id 1566721
- Location of start within SEQ ID NO 372: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 189
- gi No. 2135627
- Description: metallothionein old gene name 'MT2' - human (fragment) >gi|187528 (M26637) metallothionein old gene name 'MT2' [Homo sapiens]
- % Identity: 72.2

- Alignment Length: 18
- Location of Alignment in SEQ ID NO 374: from 27 to 44

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 375
- Ceres seq_id 1566722
- Location of start within SEQ ID NO 372: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 190
- gi No. 2135627
- Description: metallothionein old gene name 'MT2' - human (fragment) >gi|187528 (M26637) metallothionein old gene name 'MT2' [Homo sapiens]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 375: from 24 to 41

Maximum Length Sequence:
related to:
Clone IDs:
98850

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 376
- Ceres seq_id 1566727

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 377
- Ceres seq_id 1566728
- Location of start within SEQ ID NO 376: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 377: from 16 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 191
- gi No. 2829920
- Description: (AC002291) Similar to 'MADS box' transcription factors [Arabidopsis thaliana]
- % Identity: 89.3
- Alignment Length: 197
- Location of Alignment in SEQ ID NO 377: from 16 to 207

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 378
- Ceres seq_id 1566729
- Location of start within SEQ ID NO 376: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 378: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 192
- gi No. 2829920

- Description: (AC002291) Similar to 'MADS box' transcription factors [Arabidopsis thaliana]
- % Identity: 89.3
- Alignment Length: 197
- Location of Alignment in SEQ ID NO 378: from 1 to 192

Maximum Length Sequence:
related to:
Clone IDs:
98979
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 379
- Ceres seq_id 1566737
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 380
- Ceres seq_id 1566738
- Location of start within SEQ ID NO 379: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 380: from 31 to 514 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 381
- Ceres seq_id 1566739
- Location of start within SEQ ID NO 379: at 424 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 381: from 1 to 428 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 382
- Ceres seq_id 1566740
- Location of start within SEQ ID NO 379: at 451 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 382: from 1 to 419 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
99023
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 383
- Ceres seq_id 1566745
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 384
- Ceres seq_id 1566746
- Location of start within SEQ ID NO 383: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Multicopper oxidase
- Location within SEQ ID NO 384: from 5 to 144 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 193
- gi No. 5263337
- Description: (AC007727) Strong similarity to gb|X96932 ascorbate oxidase-related protein PS60 from Nicotiana tabacum and is a member of the PF|00394 Multicopper oxidase family. This gene is cut off. [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 384: from 191 to 227

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 385
- Ceres seq_id 1566747
- Location of start within SEQ ID NO 383: at 816 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 386
- Ceres seq_id 1566748
- Location of start within SEQ ID NO 383: at 1104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
99044
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 387
- Ceres seq_id 1566749
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 388
- Ceres seq_id 1566750
- Location of start within SEQ ID NO 387: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN-dependent dehydrogenase
- Location within SEQ ID NO 388: from 14 to 313 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 389
- Ceres seq_id 1566751
- Location of start within SEQ ID NO 387: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN-dependent dehydrogenase

- Location within SEQ ID NO 389: from 1 to 293 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 390
- Ceres seq_id 1566752
- Location of start within SEQ ID NO 387: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN-dependent dehydrogenase
- Location within SEQ ID NO 390: from 1 to 267 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
99235
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 391
- Ceres seq_id 1566763
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 392
- Ceres seq_id 1566764
- Location of start within SEQ ID NO 391: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 194
- gi No. 5231113
- Description: (AF141202) EIN2 [Arabidopsis thaliana]
>gi|5231115|gb|AAD41077.1|AF141203_1 (AF141203) EIN2 [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 322
- Location of Alignment in SEQ ID NO 392: from 1 to 322

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 393
- Ceres seq_id 1566765
- Location of start within SEQ ID NO 391: at 419 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 195
- gi No. 5231113
- Description: (AF141202) EIN2 [Arabidopsis thaliana]
>gi|5231115|gb|AAD41077.1|AF141203_1 (AF141203) EIN2 [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 322
- Location of Alignment in SEQ ID NO 393: from 1 to 183

Maximum Length Sequence:
related to:
Clone IDs:
99319
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 394

- Ceres seq_id 1566766

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 395
- Ceres seq_id 1566767
- Location of start within SEQ ID NO 394: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 395: from 121 to 161 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 396
- Ceres seq_id 1566768
- Location of start within SEQ ID NO 394: at 315 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 396: from 58 to 98 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 397
- Ceres seq_id 1566769
- Location of start within SEQ ID NO 394: at 357 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 397: from 44 to 84 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
99625
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 398
- Ceres seq_id 1566785

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 399
- Ceres seq_id 1566786
- Location of start within SEQ ID NO 398: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic aspartyl protease
- Location within SEQ ID NO 399: from 148 to 435 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 400
- Ceres seq_id 1566787
- Location of start within SEQ ID NO 398: at 243 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic aspartyl protease
- Location within SEQ ID NO 400: from 68 to 355 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 401
- Ceres seq_id 1566788
- Location of start within SEQ ID NO 398: at 294 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic aspartyl protease
- Location within SEQ ID NO 401: from 51 to 338 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
99836
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 402
- Ceres seq_id 1566797

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 403
- Ceres seq_id 1566798
- Location of start within SEQ ID NO 402: at 239 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 403: from 1 to 465 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 196
- gi No. 4850397
- Description: (AC007357) Strong similarity to gi|3313615 F21J9.9 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family.
- % Identity: 74.7
- Alignment Length: 529
- Location of Alignment in SEQ ID NO 403: from 1 to 477

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 404
- Ceres seq_id 1566799
- Location of start within SEQ ID NO 402: at 326 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 404: from 1 to 436 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 197
- gi No. 4850397

- Description: (AC007357) Strong similarity to gi|3313615 F21J9.9 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family.
- % Identity: 74.7
- Alignment Length: 529
- Location of Alignment in SEQ ID NO 404: from 1 to 448

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 405
- Ceres seq_id 1566800
- Location of start within SEQ ID NO 402: at 356 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 405: from 1 to 426 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 198
- gi No. 4850397
- Description: (AC007357) Strong similarity to gi|3313615 F21J9.9 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family.
- % Identity: 74.7
- Alignment Length: 529
- Location of Alignment in SEQ ID NO 405: from 1 to 438

Maximum Length Sequence:
related to:
Clone IDs:
1126
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 406
- Ceres seq_id 1566823

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 407
- Ceres seq_id 1566824
- Location of start within SEQ ID NO 406: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 407: from 29 to 347 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 408
- Ceres seq_id 1566825
- Location of start within SEQ ID NO 406: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 408: from 1 to 269 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 409
- Ceres seq_id 1566826

- Location of start within SEQ ID NO 406: at 334 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 409: from 1 to 258 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
100283
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 410
- Ceres seq_id 1566827
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 411
- Ceres seq_id 1566828
- Location of start within SEQ ID NO 410: at 322 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 199
- gi No. 2252832
- Description: (AF013293) contains similarity to S. cerevisiae ADR1 gene (PID:g924931) [Arabidopsis thaliana]
- % Identity: 72.8
- Alignment Length: 247
- Location of Alignment in SEQ ID NO 411: from 1 to 199

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 412
- Ceres seq_id 1566829
- Location of start within SEQ ID NO 410: at 643 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 200
- gi No. 2252832
- Description: (AF013293) contains similarity to S. cerevisiae ADR1 gene (PID:g924931) [Arabidopsis thaliana]
- % Identity: 72.8
- Alignment Length: 247
- Location of Alignment in SEQ ID NO 412: from 1 to 92

Maximum Length Sequence:
related to:
Clone IDs:
100412
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 413
- Ceres seq_id 1566838
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 414
- Ceres seq_id 1566839
- Location of start within SEQ ID NO 413: at 664 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 414: from 241 to 283 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 201
- gi No. 3142289
- Description: (AC002411) Strong similarity to beta-keto-Coa synthase gb|U37088 from Simmondsia chinensis. [Arabidopsis thaliana]
- % Identity: 93.1
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 414: from 1 to 352

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 415
- Ceres seq_id 1566840
- Location of start within SEQ ID NO 413: at 697 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 415: from 230 to 272 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 202
- gi No. 3142289
- Description: (AC002411) Strong similarity to beta-keto-Coa synthase gb|U37088 from Simmondsia chinensis. [Arabidopsis thaliana]
- % Identity: 93.1
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 415: from 1 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 416
- Ceres seq_id 1566841
- Location of start within SEQ ID NO 413: at 808 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 416: from 193 to 235 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 203
- gi No. 3142289
- Description: (AC002411) Strong similarity to beta-keto-Coa synthase gb|U37088 from Simmondsia chinensis. [Arabidopsis thaliana]
- % Identity: 93.1
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 416: from 1 to 304

Maximum Length Sequence:
related to:
Clone IDs:
11470
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 417
- Ceres seq_id 1566909
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 418

- Ceres seq_id 1566910
- Location of start within SEQ ID NO 417: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AN1-like Zinc finger
- Location within SEQ ID NO 418: from 114 to 150 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 419
- Ceres seq_id 1566911
- Location of start within SEQ ID NO 417: at 223 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AN1-like Zinc finger
- Location within SEQ ID NO 419: from 84 to 120 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 420
- Ceres seq_id 1566912
- Location of start within SEQ ID NO 417: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AN1-like Zinc finger
- Location within SEQ ID NO 420: from 82 to 118 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
102343
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 421
- Ceres seq_id 1566913

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 422
- Ceres seq_id 1566914
- Location of start within SEQ ID NO 421: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 204
- gi No. 4982497
- Description: (AC000107) F17F8.3 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 422: from 1 to 120

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 423
- Ceres seq_id 1566915
- Location of start within SEQ ID NO 421: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 205
- gi No. 4982497
- Description: (AC000107) F17F8.3 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 423: from 1 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 424
- Ceres seq_id 1566916
- Location of start within SEQ ID NO 421: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 206
- gi No. 4982497
- Description: (AC000107) F17F8.3 [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 424: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
102705

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 425
- Ceres seq_id 1566917

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 426
- Ceres seq_id 1566918
- Location of start within SEQ ID NO 425: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 426: from 165 to 218 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 427
- Ceres seq_id 1566919
- Location of start within SEQ ID NO 425: at 406 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 427: from 73 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 428
- Ceres seq_id 1566920
- Location of start within SEQ ID NO 425: at 454 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 428: from 57 to 110 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
102813
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 429
- Ceres seq_id 1566921
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 430
- Ceres seq_id 1566922
- Location of start within SEQ ID NO 429: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 430: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11575
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 431
- Ceres seq_id 1566936
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 432
- Ceres seq_id 1566937
- Location of start within SEQ ID NO 431: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 207
- gi No. 1914851
- Description: (U92454) WW domain binding protein 5; WBP5 [Mus musculus]
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 432: from 72 to 88

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 433
- Ceres seq_id 1566938
- Location of start within SEQ ID NO 431: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 208
- gi No. 1914851
- Description: (U92454) WW domain binding protein 5; WBP5 [Mus musculus]

- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 433: from 67 to 83

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 434
- Ceres seq_id 1566939
- Location of start within SEQ ID NO 431: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 209
- gi No. 1914851
- Description: (U92454) WW domain binding protein 5; WBP5 [Mus musculus]
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 434: from 57 to 73

Maximum Length Sequence:
related to:
Clone IDs:
103708
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 435
- Ceres seq_id 1566946
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 436
- Ceres seq_id 1566947
- Location of start within SEQ ID NO 435: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 436: from 23 to 112 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
103847
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 437
- Ceres seq_id 1566955
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 438
- Ceres seq_id 1566956
- Location of start within SEQ ID NO 437: at 899 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 438: from 316 to 577 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 210
- gi No. 2558654
- Description: (AC002354) No definition line found [Arabidopsis thaliana]

- % Identity: 80.4
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 438: from 22 to 603

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 439
- Ceres seq_id 1566957
- Location of start within SEQ ID NO 437: at 1400 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 439: from 149 to 410 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 211
- gi No. 2558654
- Description: (AC002354) No definition line found [Arabidopsis thaliana]
- % Identity: 80.4
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 439: from 1 to 436

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 440
- Ceres seq_id 1566958
- Location of start within SEQ ID NO 437: at 1535 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 440: from 104 to 365 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 212
- gi No. 2558654
- Description: (AC002354) No definition line found [Arabidopsis thaliana]
- % Identity: 80.4
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 440: from 1 to 391

Maximum Length Sequence:
related to:
Clone IDs:
103956
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 441
- Ceres seq_id 1566971

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 442
- Ceres seq_id 1566972
- Location of start within SEQ ID NO 441: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Actin
- Location within SEQ ID NO 442: from 5 to 254 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 443
- Ceres seq_id 1566973
- Location of start within SEQ ID NO 441: at 688 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Actin
- Location within SEQ ID NO 443: from 89 to 216 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
104182
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 444
- Ceres seq_id 1566985
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 445
- Ceres seq_id 1566986
- Location of start within SEQ ID NO 444: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 3' exoribonuclease family
- Location within SEQ ID NO 445: from 28 to 262 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 446
- Ceres seq_id 1566987
- Location of start within SEQ ID NO 444: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 3' exoribonuclease family
- Location within SEQ ID NO 446: from 21 to 255 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 447
- Ceres seq_id 1566988
- Location of start within SEQ ID NO 444: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 3' exoribonuclease family
- Location within SEQ ID NO 447: from 1 to 202 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
104261
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 448

- Ceres seq_id 1566989

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 449
- Ceres seq_id 1566990
- Location of start within SEQ ID NO 448: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zn-finger in Ran binding protein and others.
- Location within SEQ ID NO 449: from 241 to 272 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 213
- gi No. 2760323
- Description: (AC002130) F1N21.8 [Arabidopsis thaliana]
- % Identity: 97.9
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 449: from 1 to 290

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 450
- Ceres seq_id 1566991
- Location of start within SEQ ID NO 448: at 249 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zn-finger in Ran binding protein and others.
- Location within SEQ ID NO 450: from 198 to 229 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 214
- gi No. 2760323
- Description: (AC002130) F1N21.8 [Arabidopsis thaliana]
- % Identity: 97.9
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 450: from 1 to 247

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 451
- Ceres seq_id 1566992
- Location of start within SEQ ID NO 448: at 309 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zn-finger in Ran binding protein and others.
- Location within SEQ ID NO 451: from 178 to 209 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 215
- gi No. 2760323
- Description: (AC002130) F1N21.8 [Arabidopsis thaliana]
- % Identity: 97.9
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 451: from 1 to 227

Maximum Length Sequence:
related to:
Clone IDs:
104283

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 452

- Ceres seq_id 1566993

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 453
- Ceres seq_id 1566994
- Location of start within SEQ ID NO 452: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Polygalacturonase (pectinase)
- Location within SEQ ID NO 453: from 202 to 320 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 454
- Ceres seq_id 1566995
- Location of start within SEQ ID NO 452: at 185 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Polygalacturonase (pectinase)
- Location within SEQ ID NO 454: from 141 to 259 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 455
- Ceres seq_id 1566996
- Location of start within SEQ ID NO 452: at 596 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Polygalacturonase (pectinase)
- Location within SEQ ID NO 455: from 4 to 122 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11704

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 456
- Ceres seq_id 1567007

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 457
- Ceres seq_id 1567008
- Location of start within SEQ ID NO 456: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 458
- Ceres seq_id 1567009
- Location of start within SEQ ID NO 456: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Heme-binding domain in cytochrome b5 and oxidoreductases

- Location within SEQ ID NO 458: from 9 to 84 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 459
- Ceres seq_id 1567010
- Location of start within SEQ ID NO 456: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Heme-binding domain in cytochrome b5 and oxidoreductases
- Location within SEQ ID NO 459: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
104778
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 460
- Ceres seq_id 1567015
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 461
- Ceres seq_id 1567016
- Location of start within SEQ ID NO 460: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 461: from 231 to 294 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 462
- Ceres seq_id 1567017
- Location of start within SEQ ID NO 460: at 194 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 462: from 167 to 230 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 463
- Ceres seq_id 1567018
- Location of start within SEQ ID NO 460: at 440 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 463: from 85 to 148 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
105347
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 464
- Ceres seq_id 1567045
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 465
- Ceres seq_id 1567046
- Location of start within SEQ ID NO 464: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 465: from 65 to 134 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 466
- Ceres seq_id 1567047
- Location of start within SEQ ID NO 464: at 902 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 466: from 1 to 139 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 467
- Ceres seq_id 1567048
- Location of start within SEQ ID NO 464: at 929 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 467: from 1 to 130 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
106102
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 468
- Ceres seq_id 1567082
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 469
- Ceres seq_id 1567083
- Location of start within SEQ ID NO 468: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 216
- gi No. 2129753
- Description: threonine synthase (EC 4.2.99.2) precursor - Arabidopsis thaliana (fragment) >gi|1448917 (L41666) threonine synthase [Arabidopsis thaliana]

- % Identity: 81.7
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 469: from 1 to 109

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 470
- Ceres seq_id 1567084
- Location of start within SEQ ID NO 468: at 45 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 217
- gi No. 2129753
- Description: threonine synthase (EC 4.2.99.2) precursor - Arabidopsis thaliana (fragment) >gi|1448917 (L41666) threonine synthase [Arabidopsis thaliana]
- % Identity: 81.7
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 470: from 1 to 95

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 471
- Ceres seq_id 1567085
- Location of start within SEQ ID NO 468: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 218
- gi No. 2129753
- Description: threonine synthase (EC 4.2.99.2) precursor - Arabidopsis thaliana (fragment) >gi|1448917 (L41666) threonine synthase [Arabidopsis thaliana]
- % Identity: 81.7
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 471: from 1 to 84

Maximum Length Sequence:
related to:
Clone IDs:
106168
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 472
- Ceres seq_id 1567088
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 473
- Ceres seq_id 1567089
- Location of start within SEQ ID NO 472: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 473: from 35 to 295 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 219
- gi No. 4100323
- Description: (U96643) receptor-like protein kinase [Helianthus annuus]
- % Identity: 81.8

- Alignment Length: 11
- Location of Alignment in SEQ ID NO 473: from 44 to 54

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 474
- Ceres seq_id 1567090
- Location of start within SEQ ID NO 472: at 314 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 474: from 1 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 220
- gi No. 2129682
- Description: protein kinase AK13 - Arabidopsis thaliana (fragment) >gi|2146750|pir||S66326 protein kinase AK13 - Arabidopsis thaliana (fragment) >gi|1261864|emb|CAA60521| (X86958) protein kinase catalytic domain (fragment) [Arabidopsis thaliana]
- % Identity: 70.9
- Alignment Length: 55
- Location of Alignment in SEQ ID NO 474: from 57 to 111

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 475
- Ceres seq_id 1567091
- Location of start within SEQ ID NO 472: at 440 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 475: from 1 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 221
- gi No. 2129682
- Description: protein kinase AK13 - Arabidopsis thaliana (fragment) >gi|2146750|pir||S66326 protein kinase AK13 - Arabidopsis thaliana (fragment) >gi|1261864|emb|CAA60521| (X86958) protein kinase catalytic domain (fragment) [Arabidopsis thaliana]
- % Identity: 70.9
- Alignment Length: 55
- Location of Alignment in SEQ ID NO 475: from 15 to 69

Maximum Length Sequence:
related to:
Clone IDs:
106227
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 476
- Ceres seq_id 1567092

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 477
- Ceres seq_id 1567093
- Location of start within SEQ ID NO 476: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 477: from 53 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 222
- gi No. 3023848
- Description: GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (WD-40 REPEAT AUXIN-DEPENDENT PROTEIN ARCA) >gi|2289095 (U77381) WD-40 repeat protein [Arabidopsis thaliana]
- % Identity: 89.1
- Alignment Length: 239
- Location of Alignment in SEQ ID NO 477: from 1 to 238

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 478
- Ceres seq_id 1567094
- Location of start within SEQ ID NO 476: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 478: from 43 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 223
- gi No. 3023848
- Description: GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (WD-40 REPEAT AUXIN-DEPENDENT PROTEIN ARCA) >gi|2289095 (U77381) WD-40 repeat protein [Arabidopsis thaliana]
- % Identity: 89.1
- Alignment Length: 239
- Location of Alignment in SEQ ID NO 478: from 1 to 228

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 479
- Ceres seq_id 1567095
- Location of start within SEQ ID NO 476: at 689 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 224
- gi No. 3023848
- Description: GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (WD-40 REPEAT AUXIN-DEPENDENT PROTEIN ARCA) >gi|2289095 (U77381) WD-40 repeat protein [Arabidopsis thaliana]
- % Identity: 79.5
- Alignment Length: 88
- Location of Alignment in SEQ ID NO 479: from 36 to 123

Maximum Length Sequence:
related to:
Clone IDs:
106335
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 480
- Ceres seq_id 1567096

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 481
- Ceres seq_id 1567097
- Location of start within SEQ ID NO 480: at 244 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 481: from 32 to 272 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 225
- gi No. 4688615
- Description: (AJ131636) D-type cyclin [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 481: from 1 to 308

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 482
- Ceres seq_id 1567098
- Location of start within SEQ ID NO 480: at 307 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 482: from 11 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 226
- gi No. 4688615
- Description: (AJ131636) D-type cyclin [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 482: from 1 to 287

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 483
- Ceres seq_id 1567099
- Location of start within SEQ ID NO 480: at 349 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 483: from 1 to 237 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 227
- gi No. 4688615
- Description: (AJ131636) D-type cyclin [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 483: from 1 to 273

Maximum Length Sequence:
related to:
Clone IDs:
106897
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 484
- Ceres seq_id 1567116
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 485
- Ceres seq_id 1567117
- Location of start within SEQ ID NO 484: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lipocalin
- Location within SEQ ID NO 485: from 143 to 216 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 486
- Ceres seq_id 1567118
- Location of start within SEQ ID NO 484: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lipocalin
- Location within SEQ ID NO 486: from 138 to 211 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 487
- Ceres seq_id 1567119
- Location of start within SEQ ID NO 484: at 378 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- lipocalin
- Location within SEQ ID NO 487: from 18 to 91 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11929
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 488
- Ceres seq_id 1567120
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 489
- Ceres seq_id 1567121
- Location of start within SEQ ID NO 488: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 489: from 30 to 265 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 228
- gi No. 4678311
- Description: (AL049655) aquaporin/MIP-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 489: from 1 to 286

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 490
- Ceres seq_id 1567122
- Location of start within SEQ ID NO 488: at 264 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 490: from 1 to 210 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 229
- gi No. 4678311
- Description: (AL049655) aquaporin/MIP-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 490: from 1 to 231

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 491
- Ceres seq_id 1567123
- Location of start within SEQ ID NO 488: at 363 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 491: from 1 to 177 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 230
- gi No. 4678311
- Description: (AL049655) aquaporin/MIP-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 491: from 1 to 198

Maximum Length Sequence:
related to:
Clone IDs:
11983
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 492
- Ceres seq_id 1567128
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 493
- Ceres seq_id 1567129
- Location of start within SEQ ID NO 492: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein
- Location within SEQ ID NO 493: from 31 to 191 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 494
- Ceres seq_id 1567130
- Location of start within SEQ ID NO 492: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein

- Location within SEQ ID NO 494: from 7 to 167 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
11988
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 495
- Ceres seq_id 1567133
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 496
- Ceres seq_id 1567134
- Location of start within SEQ ID NO 495: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 496: from 35 to 355 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 497
- Ceres seq_id 1567135
- Location of start within SEQ ID NO 495: at 25 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 497: from 27 to 347 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 498
- Ceres seq_id 1567136
- Location of start within SEQ ID NO 495: at 286 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 498: from 1 to 260 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
107817
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 499
- Ceres seq_id 1567149
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 500
- Ceres seq_id 1567150
- Location of start within SEQ ID NO 499: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 231
- gi No. 1707022
- Description: (U78721) proline-rich protein isolog [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 500: from 10 to 37

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 501
- Ceres seq_id 1567151
- Location of start within SEQ ID NO 499: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 232
- gi No. 1707022
- Description: (U78721) proline-rich protein isolog [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 501: from 1 to 28

Maximum Length Sequence:
related to:
Clone IDs:
108032
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 502
- Ceres seq_id 1567164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 503
- Ceres seq_id 1567165
- Location of start within SEQ ID NO 502: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ank repeat
- Location within SEQ ID NO 503: from 269 to 301 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 233
- gi No. 584741
- Description: ANKYRIN REPEAT PROTEIN (AKRP) >gi|322461|pir||JQ1729 ankyrin-repeat protein - Arabidopsis thaliana >gi|166744 (M82883) ankyrin repeat-containing protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 407
- Location of Alignment in SEQ ID NO 503: from 1 to 407

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 504
- Ceres seq_id 1567166
- Location of start within SEQ ID NO 502: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ank repeat
- Location within SEQ ID NO 504: from 259 to 291 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 234
- gi No. 584741
- Description: ANKYRIN REPEAT PROTEIN (AKRP) >gi|322461|pir||JQ1729 ankyrin-repeat protein - Arabidopsis thaliana >gi|166744 (M82883) ankyrin repeat-containing protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 407
- Location of Alignment in SEQ ID NO 504: from 1 to 397

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 505
- Ceres seq_id 1567167
- Location of start within SEQ ID NO 502: at 422 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ank repeat
- Location within SEQ ID NO 505: from 129 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 235
- gi No. 584741
- Description: ANKYRIN REPEAT PROTEIN (AKRP) >gi|322461|pir||JQ1729 ankyrin-repeat protein - Arabidopsis thaliana >gi|166744 (M82883) ankyrin repeat-containing protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 407
- Location of Alignment in SEQ ID NO 505: from 1 to 267

Maximum Length Sequence:
related to:
Clone IDs:
108050
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 506
- Ceres seq_id 1567168

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 507
- Ceres seq_id 1567169
- Location of start within SEQ ID NO 506: at 465 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DedA family
- Location within SEQ ID NO 507: from 1 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 508
- Ceres seq_id 1567170
- Location of start within SEQ ID NO 506: at 471 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DedA family
- Location within SEQ ID NO 508: from 1 to 112 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 509
- Ceres seq_id 1567171
- Location of start within SEQ ID NO 506: at 579 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DedA family
- Location within SEQ ID NO 509: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
108109
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 510
- Ceres seq_id 1567172
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 511
- Ceres seq_id 1567173
- Location of start within SEQ ID NO 510: at 152 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 236
- gi No. 3618318
- Description: (AB001887) zinc finger protein [Oryza sativa]
- % Identity: 70.5
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 511: from 358 to 401

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 512
- Ceres seq_id 1567174
- Location of start within SEQ ID NO 510: at 155 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 237
- gi No. 3618318
- Description: (AB001887) zinc finger protein [Oryza sativa]
- % Identity: 70.5
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 512: from 357 to 400

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 513
- Ceres seq_id 1567175
- Location of start within SEQ ID NO 510: at 443 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 238
- gi No. 3618318
- Description: (AB001887) zinc finger protein [Oryza sativa]

- % Identity: 70.5
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 513: from 261 to 304

Maximum Length Sequence:
related to:
Clone IDs:
108174
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 514
- Ceres seq_id 1567183
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 515
- Ceres seq_id 1567184
- Location of start within SEQ ID NO 514: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 515: from 22 to 297 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 239
- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 515: from 1 to 305

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 516
- Ceres seq_id 1567185
- Location of start within SEQ ID NO 514: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 516: from 1 to 274 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 240
- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 516: from 1 to 282

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 517
- Ceres seq_id 1567186
- Location of start within SEQ ID NO 514: at 149 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 517: from 1 to 258 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 241

- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 517: from 1 to 266

Maximum Length Sequence:
related to:
Clone IDs:
108568
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 518
- Ceres seq_id 1567209
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 519
- Ceres seq_id 1567210
- Location of start within SEQ ID NO 518: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 519: from 323 to 363 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 520
- Ceres seq_id 1567211
- Location of start within SEQ ID NO 518: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 520: from 288 to 328 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 521
- Ceres seq_id 1567212
- Location of start within SEQ ID NO 518: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 521: from 286 to 326 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
108583
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 522
- Ceres seq_id 1567213
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 523
- Ceres seq_id 1567214
- Location of start within SEQ ID NO 522: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 523: from 12 to 116 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 242
- gi No. 4325345
- Description: (AF128393) similar to thioredoxin-like proteins (Pfam: PF00085, Score=42.9, E=1.4e-11, N=1); contains similarity to dihydroorotases (Pfam: PF00744, Score=154.9, E=1.4e-42, N=1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 382
- Location of Alignment in SEQ ID NO 523: from 10 to 391

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 524
- Ceres seq_id 1567215
- Location of start within SEQ ID NO 522: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 524: from 3 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 243
- gi No. 4325345
- Description: (AF128393) similar to thioredoxin-like proteins (Pfam: PF00085, Score=42.9, E=1.4e-11, N=1); contains similarity to dihydroorotases (Pfam: PF00744, Score=154.9, E=1.4e-42, N=1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 382
- Location of Alignment in SEQ ID NO 524: from 1 to 382

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 525
- Ceres seq_id 1567216
- Location of start within SEQ ID NO 522: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 525: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 244
- gi No. 4325345
- Description: (AF128393) similar to thioredoxin-like proteins (Pfam: PF00085, Score=42.9, E=1.4e-11, N=1); contains similarity to dihydroorotases (Pfam: PF00744, Score=154.9, E=1.4e-42, N=1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 382
- Location of Alignment in SEQ ID NO 525: from 1 to 344

Maximum Length Sequence:
related to:

Clone IDs:
108949
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 526
- Ceres seq_id 1567240
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 527
- Ceres seq_id 1567241
- Location of start within SEQ ID NO 526: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 245
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]
- % Identity: 70.2
- Alignment Length: 151
- Location of Alignment in SEQ ID NO 527: from 1 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 528
- Ceres seq_id 1567242
- Location of start within SEQ ID NO 526: at 330 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 246
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]
- % Identity: 70.2
- Alignment Length: 151
- Location of Alignment in SEQ ID NO 528: from 1 to 84

Maximum Length Sequence:
related to:
Clone IDs:
109103
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 529
- Ceres seq_id 1567251
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 530
- Ceres seq_id 1567252
- Location of start within SEQ ID NO 529: at 303 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 531
- Ceres seq_id 1567253
- Location of start within SEQ ID NO 529: at 608 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 247
- gi No. 3873550
- Description: (AL033534) serine-rich protein [Schizosaccharomyces pombe]
- % Identity: 71.4
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 531: from 150 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 532
- Ceres seq_id 1567254
- Location of start within SEQ ID NO 529: at 917 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 248
- gi No. 3873550
- Description: (AL033534) serine-rich protein [Schizosaccharomyces pombe]
- % Identity: 71.4
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 532: from 47 to 72

Maximum Length Sequence:
related to:
Clone IDs:
109368

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 533
- Ceres seq_id 1567261

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 534
- Ceres seq_id 1567262
- Location of start within SEQ ID NO 533: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sodium:sulfate symporter transmembrane region
- Location within SEQ ID NO 534: from 1 to 378 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 249
- gi No. 2499535
- Description: 2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR >gi|595681 (U13238) 2-oxoglutarate/malate translocator [Spinacia oleracea]
- % Identity: 88.6
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 534: from 1 to 378

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 535
- Ceres seq_id 1567263
- Location of start within SEQ ID NO 533: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sodium:sulfate symporter transmembrane region
- Location within SEQ ID NO 535: from 1 to 329 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 250
- gi No. 2499535
- Description: 2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR >gi|595681 (U13238) 2-oxoglutarate/malate translocator [Spinacia oleracea]
- % Identity: 88.6
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 535: from 1 to 329

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 536
- Ceres seq_id 1567264
- Location of start within SEQ ID NO 533: at 496 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sodium:sulfate symporter transmembrane region
- Location within SEQ ID NO 536: from 1 to 213 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 251
- gi No. 2499535
- Description: 2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR >gi|595681 (U13238) 2-oxoglutarate/malate translocator [Spinacia oleracea]
- % Identity: 88.6
- Alignment Length: 378
- Location of Alignment in SEQ ID NO 536: from 1 to 213

Maximum Length Sequence:
related to:
Clone IDs:
109398
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 537
- Ceres seq_id 1567265

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 538
- Ceres seq_id 1567266
- Location of start within SEQ ID NO 537: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 538: from 188 to 236 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 539
- Ceres seq_id 1567267
- Location of start within SEQ ID NO 537: at 191 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 539: from 125 to 173 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 540
- Ceres seq_id 1567268

- Location of start within SEQ ID NO 537: at 254 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 540: from 104 to 152 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
109432
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 541
- Ceres seq_id 1567273
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 542
- Ceres seq_id 1567274
- Location of start within SEQ ID NO 541: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 542: from 9 to 53 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 252
- gi No. 102425
- Description: collagen-related protein 2 - Hydra magnipapillata (fragment) >gi|102429|pir||S21930 mini-collagen - Hydra sp >gi|9449|emb|CAA43380| (X61046) mini-collagen [Hydra sp.]
- % Identity: 76
- Alignment Length: 25
- Location of Alignment in SEQ ID NO 542: from 20 to 44

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 543
- Ceres seq_id 1567275
- Location of start within SEQ ID NO 541: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
109598
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 544
- Ceres seq_id 1567284
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 545
- Ceres seq_id 1567285
- Location of start within SEQ ID NO 544: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 545: from 170 to 230 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 253
- gi No. 1173309
- Description: MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S10 PRECURSOR >gi|2119078|pir||S66361 ribosomal protein S10 precursor - Arabidopsis thaliana >gi|517331|emb|CAA56711| (X80694) ribosomal protein S10 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 241
- Location of Alignment in SEQ ID NO 545: from 15 to 255

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 546
- Ceres seq_id 1567286
- Location of start within SEQ ID NO 544: at 44 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 546: from 156 to 216 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 254
- gi No. 1173309
- Description: MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S10 PRECURSOR >gi|2119078|pir||S66361 ribosomal protein S10 precursor - Arabidopsis thaliana >gi|517331|emb|CAA56711| (X80694) ribosomal protein S10 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 241
- Location of Alignment in SEQ ID NO 546: from 1 to 241

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 547
- Ceres seq_id 1567287
- Location of start within SEQ ID NO 544: at 248 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 547: from 88 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 255
- gi No. 1173309
- Description: MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S10 PRECURSOR >gi|2119078|pir||S66361 ribosomal protein S10 precursor - Arabidopsis thaliana >gi|517331|emb|CAA56711| (X80694) ribosomal protein S10 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 241
- Location of Alignment in SEQ ID NO 547: from 1 to 173

Maximum Length Sequence:
related to:
Clone IDs:
1222
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 548
- Ceres seq_id 1567288

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 549
- Ceres seq_id 1567289
- Location of start within SEQ ID NO 548: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 549: from 164 to 289 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 256
- gi No. 1172873
- Description: CYSTEINE PROTEINASE RD21A PRECURSOR >gi|541857|pir||JN0719 drought-inducible cysteine proteinase (EC 3.4.22.-) RD21A precursor - Arabidopsis thaliana >gi|435619|dbj|BAA02374| (D13043) thiol protease [Arabidopsis thaliana]
- % Identity: 88.3
- Alignment Length: 300
- Location of Alignment in SEQ ID NO 549: from 28 to 295

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 550
- Ceres seq_id 1567290
- Location of start within SEQ ID NO 548: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 550: from 137 to 262 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 257
- gi No. 1172873
- Description: CYSTEINE PROTEINASE RD21A PRECURSOR >gi|541857|pir||JN0719 drought-inducible cysteine proteinase (EC 3.4.22.-) RD21A precursor - Arabidopsis thaliana >gi|435619|dbj|BAA02374| (D13043) thiol protease [Arabidopsis thaliana]
- % Identity: 88.3
- Alignment Length: 300
- Location of Alignment in SEQ ID NO 550: from 1 to 268

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 551
- Ceres seq_id 1567291
- Location of start within SEQ ID NO 548: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 551: from 130 to 255 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 258
- gi No. 1172873
- Description: CYSTEINE PROTEINASE RD21A PRECURSOR >gi|541857|pir||JN0719 drought-inducible cysteine proteinase (EC 3.4.22.-) RD21A precursor - Arabidopsis thaliana >gi|435619|dbj|BAA02374| (D13043) thiol protease [Arabidopsis thaliana]
- % Identity: 88.3
- Alignment Length: 300

- Location of Alignment in SEQ ID NO 551: from 1 to 261

Maximum Length Sequence:
related to:
Clone IDs:
110419
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 552
- Ceres seq_id 1567305
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 553
- Ceres seq_id 1567306
- Location of start within SEQ ID NO 552: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 553: from 189 to 229 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 554
- Ceres seq_id 1567307
- Location of start within SEQ ID NO 552: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 554: from 164 to 204 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 555
- Ceres seq_id 1567308
- Location of start within SEQ ID NO 552: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 555: from 153 to 193 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
110454
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 556
- Ceres seq_id 1567309
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 557
- Ceres seq_id 1567310
- Location of start within SEQ ID NO 556: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GATA zinc finger
- Location within SEQ ID NO 557: from 229 to 267 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 259
- gi No. 4885255
- Description: ref|NP_005248.1|pGATA6| GATA-binding protein 6 >gi|2494684|sp|Q92908|GAT6_HUMAN TRANSCRIPTION FACTOR GATA-6 (GATA BINDING FACTOR-6) >gi|1655915 (U66075) hGATA-6 [Homo sapiens] >gi|2506076|dbj|BAA22621| (D87811) GATA-6 [Homo sapiens]
- % Identity: 83.3
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 557: from 33 to 44

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 558
- Ceres seq_id 1567311
- Location of start within SEQ ID NO 556: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GATA zinc finger
- Location within SEQ ID NO 558: from 136 to 174 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 559
- Ceres seq_id 1567312
- Location of start within SEQ ID NO 556: at 435 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GATA zinc finger
- Location within SEQ ID NO 559: from 123 to 161 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
12301
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 560
- Ceres seq_id 1567317

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 561
- Ceres seq_id 1567318
- Location of start within SEQ ID NO 560: at 44 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 561: from 24 to 217 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 260
- gi No. 2827712
- Description: (AL021684) endoxyloglucan tranferase-like protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 561: from 24 to 244

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 562
- Ceres seq_id 1567319
- Location of start within SEQ ID NO 560: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 562: from 1 to 194 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 261
- gi No. 2827712
- Description: (AL021684) endoxyloglucan tranferase-like protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 562: from 1 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 563
- Ceres seq_id 1567320
- Location of start within SEQ ID NO 560: at 194 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 563: from 1 to 167 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 262
- gi No. 2827712
- Description: (AL021684) endoxyloglucan tranferase-like protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 563: from 1 to 194

Maximum Length Sequence:
related to:
Clone IDs:
110908
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 564
- Ceres seq_id 1567325

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 565
- Ceres seq_id 1567326
- Location of start within SEQ ID NO 564: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 263
- gi No. 854743
- Description: (U28888) neurogenic differentiation factor [Mus musculus]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 565: from 11 to 31

Maximum Length Sequence:
related to:
Clone IDs:
111056
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 566
- Ceres seq_id 1567331
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 567
- Ceres seq_id 1567332
- Location of start within SEQ ID NO 566: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxylesterases
- Location within SEQ ID NO 567: from 74 to 174 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 264
- gi No. 5668813
- Description: (AC007519) Similar to gb|X77136 HSR203J protein from Nicotiana tabacum and is a member of the PF|00135 Carboxylesterase family. ESTs gb|Z25688 and gb|F14025 come from this gene. [Arabidopsis thaliana]
- % Identity: 85.3
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 567: from 15 to 223

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 568
- Ceres seq_id 1567333
- Location of start within SEQ ID NO 566: at 45 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxylesterases
- Location within SEQ ID NO 568: from 60 to 160 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 265
- gi No. 5668813
- Description: (AC007519) Similar to gb|X77136 HSR203J protein from Nicotiana tabacum and is a member of the PF|00135 Carboxylesterase family. ESTs gb|Z25688 and gb|F14025 come from this gene. [Arabidopsis thaliana]
- % Identity: 85.3
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 568: from 1 to 209

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 569
- Ceres seq_id 1567334
- Location of start within SEQ ID NO 566: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxylesterases
- Location within SEQ ID NO 569: from 1 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 266
- gi No. 5668813

- Description: (AC007519) Similar to gb|X77136 HSR203J protein from Nicotiana tabacum and is a member of the PF|00135 Carboxylesterase family. ESTs gb|Z25688 and gb|F14025 come from this gene. [Arabidopsis thaliana]
  - % Identity: 85.3
  - Alignment Length: 211
  - Location of Alignment in SEQ ID NO 569: from 1 to 135

Maximum Length Sequence:
related to:
Clone IDs:
111254
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 570
- Ceres seq_id 1567335

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 571
- Ceres seq_id 1567336
- Location of start within SEQ ID NO 570: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 571: from 196 to 513 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 572
- Ceres seq_id 1567337
- Location of start within SEQ ID NO 570: at 275 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 572: from 148 to 465 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 573
- Ceres seq_id 1567338
- Location of start within SEQ ID NO 570: at 368 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 573: from 117 to 434 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
111312
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 574
- Ceres seq_id 1567343

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 575
- Ceres seq_id 1567344
- Location of start within SEQ ID NO 574: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 575: from 235 to 296 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 576
- Ceres seq_id 1567345
- Location of start within SEQ ID NO 574: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 576: from 207 to 268 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 577
- Ceres seq_id 1567346
- Location of start within SEQ ID NO 574: at 415 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 577: from 97 to 158 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
111427
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 578
- Ceres seq_id 1567347

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 579
- Ceres seq_id 1567348
- Location of start within SEQ ID NO 578: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 579: from 30 to 190 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 267
- gi No. 2275200
- Description: (AC002337) polyribonucleotide phosphorylase isolog [Arabidopsis thaliana]
- % Identity: 73.6
- Alignment Length: 53
- Location of Alignment in SEQ ID NO 579: from 39 to 91

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 580
- Ceres seq_id 1567349

- Location of start within SEQ ID NO 578: at 409 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 580: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
111694
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 581
- Ceres seq_id 1567371
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 582
- Ceres seq_id 1567372
- Location of start within SEQ ID NO 581: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- BTB/POZ domain
- Location within SEQ ID NO 582: from 11 to 128 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 583
- Ceres seq_id 1567373
- Location of start within SEQ ID NO 581: at 370 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 584
- Ceres seq_id 1567374
- Location of start within SEQ ID NO 581: at 415 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
112003
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 585
- Ceres seq_id 1567383
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 586
- Ceres seq_id 1567384
- Location of start within SEQ ID NO 585: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)

- Location within SEQ ID NO 586: from 177 to 218 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 587
- Ceres seq_id 1567385
- Location of start within SEQ ID NO 585: at 512 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 587: from 48 to 89 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
112346
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 588
- Ceres seq_id 1567390
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 589
- Ceres seq_id 1567391
- Location of start within SEQ ID NO 588: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 589: from 225 to 575 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 268
- gi No. 1172538
- Description: PROBABLE PECTINESTERASE PRECURSOR (PECTIN METHYLESTERASE) (PE) >gi|99823|pir||S14952 pectinesterase homolog - rape >gi|17784|emb|CAA39658| (X56195) Bp19 [Brassica napus]
- % Identity: 82.8
- Alignment Length: 589
- Location of Alignment in SEQ ID NO 589: from 1 to 588

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 590
- Ceres seq_id 1567392
- Location of start within SEQ ID NO 588: at 416 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 590: from 144 to 494 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 269
- gi No. 1172538
- Description: PROBABLE PECTINESTERASE PRECURSOR (PECTIN METHYLESTERASE) (PE) >gi|99823|pir||S14952 pectinesterase homolog - rape >gi|17784|emb|CAA39658| (X56195) Bp19 [Brassica napus]
- % Identity: 82.8
- Alignment Length: 589

- Location of Alignment in SEQ ID NO 590: from 1 to 507

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 591
- Ceres seq_id 1567393
- Location of start within SEQ ID NO 588: at 476 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 591: from 124 to 474 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 270
- gi No. 1172538
- Description: PROBABLE PECTINESTERASE PRECURSOR (PECTIN METHYLESTERASE) (PE) >gi|99823|pir||S14952 pectinesterase homolog - rape >gi|17784|emb|CAA39658| (X56195) Bp19 [Brassica napus]
- % Identity: 82.8
- Alignment Length: 589
- Location of Alignment in SEQ ID NO 591: from 1 to 487

Maximum Length Sequence:
related to:
Clone IDs:
112567
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 592
- Ceres seq_id 1567394

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 593
- Ceres seq_id 1567395
- Location of start within SEQ ID NO 592: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 593: from 68 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 271
- gi No. 3953471
- Description: (AC002328) F2202.16 [Arabidopsis thaliana]
- % Identity: 92.2
- Alignment Length: 206
- Location of Alignment in SEQ ID NO 593: from 1 to 206

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 594
- Ceres seq_id 1567396
- Location of start within SEQ ID NO 592: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 594: from 45 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 272
- gi No. 3953471
- Description: (AC002328) F2202.16 [Arabidopsis thaliana]

- % Identity: 92.2
- Alignment Length: 206
- Location of Alignment in SEQ ID NO 594: from 1 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 595
- Ceres seq_id 1567397
- Location of start within SEQ ID NO 592: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
112834
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 596
- Ceres seq_id 1567405
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 597
- Ceres seq_id 1567406
- Location of start within SEQ ID NO 596: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 597: from 178 to 416 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 598
- Ceres seq_id 1567407
- Location of start within SEQ ID NO 596: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 598: from 99 to 337 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 599
- Ceres seq_id 1567408
- Location of start within SEQ ID NO 596: at 522 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 599: from 20 to 258 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
112853
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 600
- Ceres seq_id 1567417

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 601
- Ceres seq_id 1567418
- Location of start within SEQ ID NO 600: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 273
- gi No. 3024665
- Description: STRICTOSIDINE SYNTHASE 3 PRECURSOR >gi|1754987 (U43946) strictosidine synthase [Arabidopsis thaliana]
- % Identity: 93.6
- Alignment Length: 329
- Location of Alignment in SEQ ID NO 601: from 1 to 325

Maximum Length Sequence:
related to:
Clone IDs:
112955

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 602
- Ceres seq_id 1567419

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 603
- Ceres seq_id 1567420
- Location of start within SEQ ID NO 602: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 274
- gi No. 2262115
- Description: (AC002343) cellulose synthase isolog [Arabidopsis thaliana]
- % Identity: 92.5
- Alignment Length: 226
- Location of Alignment in SEQ ID NO 603: from 1 to 217

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 604
- Ceres seq_id 1567421
- Location of start within SEQ ID NO 602: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 275
- gi No. 2262115
- Description: (AC002343) cellulose synthase isolog [Arabidopsis thaliana]
- % Identity: 92.5
- Alignment Length: 226
- Location of Alignment in SEQ ID NO 604: from 1 to 180

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 605
- Ceres seq_id 1567422
- Location of start within SEQ ID NO 602: at 697 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 276
- gi No. 2262115
- Description: (AC002343) cellulose synthase isolog [Arabidopsis thaliana]
- % Identity: 93.8
- Alignment Length: 177
- Location of Alignment in SEQ ID NO 605: from 7 to 183

Maximum Length Sequence:
related to:
Clone IDs:
113439
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 606
- Ceres seq_id 1567445
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 607
- Ceres seq_id 1567446
- Location of start within SEQ ID NO 606: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 607: from 13 to 219 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 277
- gi No. 4204297
- Description: (AC003027) ADK1 [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 478
- Location of Alignment in SEQ ID NO 607: from 5 to 479

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 608
- Ceres seq_id 1567447
- Location of start within SEQ ID NO 606: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 608: from 9 to 215 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 278
- gi No. 4204297
- Description: (AC003027) ADK1 [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 478
- Location of Alignment in SEQ ID NO 608: from 1 to 475

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 609
- Ceres seq_id 1567448
- Location of start within SEQ ID NO 606: at 441 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 609: from 1 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 279
- gi No. 4204297
- Description: (AC003027) ADK1 [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 478
- Location of Alignment in SEQ ID NO 609: from 1 to 419

Maximum Length Sequence:
related to:
Clone IDs:
12604
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 610
- Ceres seq_id 1567449

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 611
- Ceres seq_id 1567450
- Location of start within SEQ ID NO 610: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 280
- gi No. 2191138
- Description: (AF007269) A_IG002N01.18 gene product [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 611: from 1 to 58

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 612
- Ceres seq_id 1567451
- Location of start within SEQ ID NO 610: at 354 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 613
- Ceres seq_id 1567452
- Location of start within SEQ ID NO 610: at 387 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
113759
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 614
- Ceres seq_id 1567453

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 615
- Ceres seq_id 1567454
- Location of start within SEQ ID NO 614: at 286 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 615: from 1 to 210 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 281
- gi No. 1389699
- Description: (U38845) nitrilase 1 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 304
- Location of Alignment in SEQ ID NO 615: from 1 to 224

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 616
- Ceres seq_id 1567455
- Location of start within SEQ ID NO 614: at 379 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 616: from 1 to 179 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 282
- gi No. 1389699
- Description: (U38845) nitrilase 1 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 304
- Location of Alignment in SEQ ID NO 616: from 1 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 617
- Ceres seq_id 1567456
- Location of start within SEQ ID NO 614: at 490 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 617: from 1 to 142 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 283
- gi No. 1389699
- Description: (U38845) nitrilase 1 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 304
- Location of Alignment in SEQ ID NO 617: from 1 to 156

Maximum Length Sequence:
related to:
Clone IDs:
113896
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 618
- Ceres seq_id 1567463

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 619
- Ceres seq_id 1567464
- Location of start within SEQ ID NO 618: at 406 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class II (D, K and N)
- Location within SEQ ID NO 619: from 80 to 367 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 620
- Ceres seq_id 1567465
- Location of start within SEQ ID NO 618: at 526 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class II (D, K and N)
- Location within SEQ ID NO 620: from 40 to 327 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 621
- Ceres seq_id 1567466
- Location of start within SEQ ID NO 618: at 883 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class II (D, K and N)
- Location within SEQ ID NO 621: from 1 to 208 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1718
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 622
- Ceres seq_id 1567489

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 623
- Ceres seq_id 1567490
- Location of start within SEQ ID NO 622: at 279 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 284
- gi No. 2244798
- Description: (Z97336) elicitor like protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 623: from 1 to 97

Maximum Length Sequence:
related to:

Clone IDs:
2118
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 624
- Ceres seq_id 1567501
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 625
- Ceres seq_id 1567502
- Location of start within SEQ ID NO 624: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 285
- gi No. 2190541
- Description: (AC001229) No definition line found [Arabidopsis thaliana]
- % Identity: 72.2
- Alignment Length: 74
- Location of Alignment in SEQ ID NO 625: from 8 to 79

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 626
- Ceres seq_id 1567503
- Location of start within SEQ ID NO 624: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 286
- gi No. 2190541
- Description: (AC001229) No definition line found [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 626: from 82 to 95

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 627
- Ceres seq_id 1567504
- Location of start within SEQ ID NO 624: at 22 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 287
- gi No. 2190541
- Description: (AC001229) No definition line found [Arabidopsis thaliana]
- % Identity: 72.2
- Alignment Length: 74
- Location of Alignment in SEQ ID NO 627: from 1 to 72

Maximum Length Sequence:
related to:
Clone IDs:
3201
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 628
- Ceres seq_id 1567513
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 629
- Ceres seq_id 1567514
- Location of start within SEQ ID NO 628: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 288
- gi No. 929786
- Description: (X07457) 11-1 gene protein (219 AA) [Plasmodium falciparum]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 629: from 23 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 630
- Ceres seq_id 1567515
- Location of start within SEQ ID NO 628: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 289
- gi No. 929786
- Description: (X07457) 11-1 gene protein (219 AA) [Plasmodium falciparum]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 630: from 12 to 28

Maximum Length Sequence:
related to:
Clone IDs:
2054
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 631
- Ceres seq_id 1567520
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 632
- Ceres seq_id 1567521
- Location of start within SEQ ID NO 631: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 632: from 3 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 290
- gi No. 4973262
- Description: (AF144390) thioredoxin-like 4 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 118
- Location of Alignment in SEQ ID NO 632: from 1 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 633
- Ceres seq_id 1567522
- Location of start within SEQ ID NO 631: at 205 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 633: from 1 to 64 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 291
- gi No. 4973262
- Description: (AF144390) thioredoxin-like 4 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 118
- Location of Alignment in SEQ ID NO 633: from 1 to 77

Maximum Length Sequence:
related to:
Clone IDs:
7360
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 634
- Ceres seq_id 1567526
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 635
- Ceres seq_id 1567527
- Location of start within SEQ ID NO 634: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 635: from 46 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 292
- gi No. 1491617
- Description: (X99952) peroxidase [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 237
- Location of Alignment in SEQ ID NO 635: from 1 to 212

Maximum Length Sequence:
related to:
Clone IDs:
4058
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 636
- Ceres seq_id 1567532
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 637
- Ceres seq_id 1567533
- Location of start within SEQ ID NO 636: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S15
- Location within SEQ ID NO 637: from 82 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 293
- gi No. 1350930
- Description: 40S RIBOSOMAL PROTEIN S13
- % Identity: 100
- Alignment Length: 150

- Location of Alignment in SEQ ID NO 637: from 2 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 638
- Ceres seq_id 1567534
- Location of start within SEQ ID NO 636: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S15
- Location within SEQ ID NO 638: from 79 to 144 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 294
- gi No. 1350930
- Description: 40S RIBOSOMAL PROTEIN S13
- % Identity: 100
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 638: from 1 to 148

Maximum Length Sequence:
related to:
Clone IDs:
8
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 639
- Ceres seq_id 1567544
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 640
- Ceres seq_id 1567545
- Location of start within SEQ ID NO 639: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 640: from 41 to 171 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 295
- gi No. 2407800
- Description: (Y12575) histone H2A.F/Z [Arabidopsis thaliana]
- % Identity: 90.4
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 640: from 40 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 641
- Ceres seq_id 1567546
- Location of start within SEQ ID NO 639: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 641: from 2 to 132 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 296
- gi No. 2407800
- Description: (Y12575) histone H2A.F/Z [Arabidopsis thaliana]
- % Identity: 90.4
- Alignment Length: 136

- Location of Alignment in SEQ ID NO 641: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
4513
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 642
- Ceres seq_id 1567547
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 643
- Ceres seq_id 1567548
- Location of start within SEQ ID NO 642: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 643: from 23 to 144 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 297
- gi No. 730536
- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95.7
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 643: from 5 to 144

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 644
- Ceres seq_id 1567549
- Location of start within SEQ ID NO 642: at 15 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 644: from 19 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 298
- gi No. 730536
- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95.7
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 644: from 1 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 645
- Ceres seq_id 1567550
- Location of start within SEQ ID NO 642: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 645: from 4 to 125 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 299
- gi No. 730536

- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95.7
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 645: from 1 to 125

Maximum Length Sequence:
related to:
Clone IDs:
4458
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 646
- Ceres seq_id 1567555

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 647
- Ceres seq_id 1567556
- Location of start within SEQ ID NO 646: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 647: from 1 to 90 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 300
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 647: from 1 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 648
- Ceres seq_id 1567557
- Location of start within SEQ ID NO 646: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 648: from 1 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 301
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 648: from 1 to 77

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 649
- Ceres seq_id 1567558
- Location of start within SEQ ID NO 646: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 302
- gi No. 2252857

- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 649: from 1 to 46

Maximum Length Sequence:
related to:
Clone IDs:
2688
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 650
- Ceres seq_id 1567563
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 651
- Ceres seq_id 1567564
- Location of start within SEQ ID NO 650: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ubiE/COQ5 methyltransferase family
- Location within SEQ ID NO 651: from 119 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 303
- gi No. 2129517
- Description: 24-sterol C-methyltransferase (EC 2.1.1.41) - Arabidopsis thaliana
- % Identity: 95
- Alignment Length: 365
- Location of Alignment in SEQ ID NO 651: from 1 to 249

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 652
- Ceres seq_id 1567565
- Location of start within SEQ ID NO 650: at 442 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ubiE/COQ5 methyltransferase family
- Location within SEQ ID NO 652: from 6 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 304
- gi No. 2129517
- Description: 24-sterol C-methyltransferase (EC 2.1.1.41) - Arabidopsis thaliana
- % Identity: 95
- Alignment Length: 365
- Location of Alignment in SEQ ID NO 652: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
17210
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 653
- Ceres seq_id 1567581
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 654
- Ceres seq_id 1567582

- Location of start within SEQ ID NO 653: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 654: from 181 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 305
- gi No. 4587549
- Description: (AC006577) Similar to gb|U55861 RNA binding protein nucleolysin (TIAR) from Mus musculus and contains several PF|00076 RNA recognition motif domains. ESTs gb|T21032 and gb|T44127 come from this gene. [Arabidopsis t...
- % Identity: 76.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 654: from 40 to 459

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 655
- Ceres seq_id 1567583
- Location of start within SEQ ID NO 653: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 655: from 141 to 211 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 306
- gi No. 4587549
- Description: (AC006577) Similar to gb|U55861 RNA binding protein nucleolysin (TIAR) from Mus musculus and contains several PF|00076 RNA recognition motif domains. ESTs gb|T21032 and gb|T44127 come from this gene. [Arabidopsis t...
- % Identity: 76.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 655: from 1 to 419

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 656
- Ceres seq_id 1567584
- Location of start within SEQ ID NO 653: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 656: from 128 to 198 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 307
- gi No. 4587549
- Description: (AC006577) Similar to gb|U55861 RNA binding protein nucleolysin (TIAR) from Mus musculus and contains several PF|00076 RNA recognition motif domains. ESTs gb|T21032 and gb|T44127 come from this gene. [Arabidopsis t...
- % Identity: 76.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 656: from 1 to 406

Maximum Length Sequence:

related to:
Clone IDs:
17415
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 657
- Ceres seq_id 1567589
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 658
- Ceres seq_id 1567590
- Location of start within SEQ ID NO 657: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S5
- Location within SEQ ID NO 658: from 100 to 237 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 308
- gi No. 1710756
- Description: 40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)
- % Identity: 71.5
- Alignment Length: 264
- Location of Alignment in SEQ ID NO 658: from 5 to 267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 659
- Ceres seq_id 1567591
- Location of start within SEQ ID NO 657: at 409 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S5
- Location within SEQ ID NO 659: from 11 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 309
- gi No. 1710756
- Description: 40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)
- % Identity: 71.5
- Alignment Length: 264
- Location of Alignment in SEQ ID NO 659: from 1 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 660
- Ceres seq_id 1567592
- Location of start within SEQ ID NO 657: at 442 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S5
- Location within SEQ ID NO 660: from 1 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 310
- gi No. 1710756
- Description: 40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)
- % Identity: 71.5
- Alignment Length: 264
- Location of Alignment in SEQ ID NO 660: from 1 to 167

Maximum Length Sequence:

related to:
Clone IDs:
14132
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 661
- Ceres seq_id 1567593
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 662
- Ceres seq_id 1567594
- Location of start within SEQ ID NO 661: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 311
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.5
- Alignment Length: 52
- Location of Alignment in SEQ ID NO 662: from 47 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 663
- Ceres seq_id 1567595
- Location of start within SEQ ID NO 661: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 312
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.5
- Alignment Length: 52
- Location of Alignment in SEQ ID NO 663: from 16 to 66

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 664
- Ceres seq_id 1567596
- Location of start within SEQ ID NO 661: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 313
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.5
- Alignment Length: 52
- Location of Alignment in SEQ ID NO 664: from 8 to 58

Maximum Length Sequence:
related to:
Clone IDs:
24826
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 665
- Ceres seq_id 1567605
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 666
- Ceres seq_id 1567606

- Location of start within SEQ ID NO 665: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 314
- gi No. 2224933
- Description: (AF004216) ethylene-insensitive3 [Arabidopsis thaliana] >gi|2224935 (AF004217) ethylene-insensitive3 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 111
- Location of Alignment in SEQ ID NO 666: from 1 to 111

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 667
- Ceres seq_id 1567607
- Location of start within SEQ ID NO 665: at 16 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 315
- gi No. 2224933
- Description: (AF004216) ethylene-insensitive3 [Arabidopsis thaliana] >gi|2224935 (AF004217) ethylene-insensitive3 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 111
- Location of Alignment in SEQ ID NO 667: from 1 to 106

Maximum Length Sequence:
related to:
Clone IDs:
1271

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 668
- Ceres seq_id 1567608

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 669
- Ceres seq_id 1567609
- Location of start within SEQ ID NO 668: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Elongation factor Tu family
- Location within SEQ ID NO 669: from 5 to 159 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 670
- Ceres seq_id 1567610
- Location of start within SEQ ID NO 668: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Elongation factor Tu family
- Location within SEQ ID NO 670: from 1 to 153 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 671
- Ceres seq_id 1567611
- Location of start within SEQ ID NO 668: at 105 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 671: from 1 to 125 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32793
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 672
- Ceres seq_id 1567622

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 673
- Ceres seq_id 1567623
- Location of start within SEQ ID NO 672: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 673: from 9 to 307 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 316
- gi No. 553035
- Description: (M14756) pal protein [Antirrhinum majus]
- % Identity: 71.9
- Alignment Length: 32
- Location of Alignment in SEQ ID NO 673: from 9 to 40

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 674
- Ceres seq_id 1567624
- Location of start within SEQ ID NO 672: at 5 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 674: from 8 to 306 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 317
- gi No. 553035
- Description: (M14756) pal protein [Antirrhinum majus]
- % Identity: 71.9
- Alignment Length: 32
- Location of Alignment in SEQ ID NO 674: from 8 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 675
- Ceres seq_id 1567625
- Location of start within SEQ ID NO 672: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 675: from 1 to 241 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1728
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 676
- Ceres seq_id 1567629
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 677
- Ceres seq_id 1567630
- Location of start within SEQ ID NO 676: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Annexin
- Location within SEQ ID NO 677: from 13 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 318
- gi No. 4959108
- Description: (AF083914) annexin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 317
- Location of Alignment in SEQ ID NO 677: from 1 to 317

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 678
- Ceres seq_id 1567631
- Location of start within SEQ ID NO 676: at 327 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 319
- gi No. 4959108
- Description: (AF083914) annexin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 317
- Location of Alignment in SEQ ID NO 678: from 1 to 240

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 679
- Ceres seq_id 1567632
- Location of start within SEQ ID NO 676: at 387 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 320
- gi No. 4959108
- Description: (AF083914) annexin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 317
- Location of Alignment in SEQ ID NO 679: from 1 to 220

Maximum Length Sequence:

related to:
Clone IDs:
5361
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 680
- Ceres seq_id 1567661
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 681
- Ceres seq_id 1567662
- Location of start within SEQ ID NO 680: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 681: from 70 to 194 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 321
- gi No. 2832664
- Description: (AL021710) pollen-specific protein - like [Arabidopsis thaliana]
- % Identity: 83.7
- Alignment Length: 202
- Location of Alignment in SEQ ID NO 681: from 3 to 204

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 682
- Ceres seq_id 1567663
- Location of start within SEQ ID NO 680: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 682: from 37 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 322
- gi No. 2832664
- Description: (AL021710) pollen-specific protein - like [Arabidopsis thaliana]
- % Identity: 83.7
- Alignment Length: 202
- Location of Alignment in SEQ ID NO 682: from 1 to 171

Maximum Length Sequence:
related to:
Clone IDs:
9552
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 683
- Ceres seq_id 1567674
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 684
- Ceres seq_id 1567675
- Location of start within SEQ ID NO 683: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HSF-type DNA-binding domain
- Location within SEQ ID NO 684: from 25 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 685
- Ceres seq_id 1567676
- Location of start within SEQ ID NO 683: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HSF-type DNA-binding domain
- Location within SEQ ID NO 685: from 5 to 94 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 686
- Ceres seq_id 1567677
- Location of start within SEQ ID NO 683: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HSF-type DNA-binding domain
- Location within SEQ ID NO 686: from 1 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
17737
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 687
- Ceres seq_id 1567687
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 688
- Ceres seq_id 1567688
- Location of start within SEQ ID NO 687: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class II (A)
- Location within SEQ ID NO 688: from 1 to 291 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 323
- gi No. 5734779
- Description: (AC007980) 78688 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 688: from 1 to 303

Maximum Length Sequence:
related to:
Clone IDs:
35284
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 689
- Ceres seq_id 1567700
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 690
- Ceres seq_id 1567701

- Location of start within SEQ ID NO 689: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GrpE
- Location within SEQ ID NO 690: from 116 to 314 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 324
- gi No. 4455201
- Description: (AL035440) grpE like protein [Arabidopsis thaliana]
- % Identity: 95.1
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 690: from 1 to 327

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 691
- Ceres seq_id 1567702
- Location of start within SEQ ID NO 689: at 278 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GrpE
- Location within SEQ ID NO 691: from 63 to 261 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 325
- gi No. 4455201
- Description: (AL035440) grpE like protein [Arabidopsis thaliana]
- % Identity: 95.1
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 691: from 1 to 274

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 692
- Ceres seq_id 1567703
- Location of start within SEQ ID NO 689: at 284 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GrpE
- Location within SEQ ID NO 692: from 61 to 259 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 326
- gi No. 4455201
- Description: (AL035440) grpE like protein [Arabidopsis thaliana]
- % Identity: 95.1
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 692: from 1 to 272

Maximum Length Sequence:
related to:
Clone IDs:
36576
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 693
- Ceres seq_id 1567704
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 694
- Ceres seq_id 1567705

- Location of start within SEQ ID NO 693: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 694: from 222 to 352 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 695
- Ceres seq_id 1567706
- Location of start within SEQ ID NO 693: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 695: from 218 to 348 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 696
- Ceres seq_id 1567707
- Location of start within SEQ ID NO 693: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 696: from 203 to 333 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1523
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 697
- Ceres seq_id 1567711

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 698
- Ceres seq_id 1567712
- Location of start within SEQ ID NO 697: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant thionins
- Location within SEQ ID NO 698: from 50 to 93 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 327
- gi No. 1181533
- Description: (L41245) thionin [Arabidopsis thaliana] >gi|1586834|prf||2204399B thionin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 698: from 26 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 699

- Ceres seq_id 1567713
- Location of start within SEQ ID NO 697: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant thionins
- Location within SEQ ID NO 699: from 25 to 68 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 328
- gi No. 1181533
- Description: (L41245) thionin [Arabidopsis thaliana] >gi|1586834|prf||2204399B thionin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 699: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 700
- Ceres seq_id 1567714
- Location of start within SEQ ID NO 697: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 329
- gi No. 1181533
- Description: (L41245) thionin [Arabidopsis thaliana] >gi|1586834|prf||2204399B thionin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 700: from 1 to 94

Maximum Length Sequence:
related to:
Clone IDs:
5698
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 701
- Ceres seq_id 1567715

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 702
- Ceres seq_id 1567716
- Location of start within SEQ ID NO 701: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 703
- Ceres seq_id 1567717
- Location of start within SEQ ID NO 701: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 703: from 11 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 330

- gi No. 2687686
- Description: (AJ000110) cystatin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 101
- Location of Alignment in SEQ ID NO 703: from 1 to 101

Maximum Length Sequence:
related to:
Clone IDs:
26943
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 704
- Ceres seq_id 1567721

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 705
- Ceres seq_id 1567722
- Location of start within SEQ ID NO 704: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 706
- Ceres seq_id 1567723
- Location of start within SEQ ID NO 704: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 706: from 3 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 331
- gi No. 461729
- Description: 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES)
>gi|2146744|pir||S65597 probable chaperonin, 10K - Arabidopsis thaliana
>gi|166662 (L02843) 10 kDa chaperonin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 706: from 1 to 98

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 707
- Ceres seq_id 1567724
- Location of start within SEQ ID NO 704: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 707: from 2 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 332
- gi No. 461729
- Description: 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES)
>gi|2146744|pir||S65597 probable chaperonin, 10K - Arabidopsis thaliana
>gi|166662 (L02843) 10 kDa chaperonin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 707: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
17432
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 708
- Ceres seq_id 1567738
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 709
- Ceres seq_id 1567739
- Location of start within SEQ ID NO 708: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase FAD/NAD-binding domain
- Location within SEQ ID NO 709: from 154 to 260 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 333
- gi No. 4240116
- Description: (AB007799) NADH-cytochrome b5 reductase [Arabidopsis thaliana] >gi|4240118|dbj|BAA74838| (AB007800) NADH-cytochrome b5 reductase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 709: from 1 to 281

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 710
- Ceres seq_id 1567740
- Location of start within SEQ ID NO 708: at 505 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase FAD/NAD-binding domain
- Location within SEQ ID NO 710: from 37 to 143 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 334
- gi No. 4240116
- Description: (AB007799) NADH-cytochrome b5 reductase [Arabidopsis thaliana] >gi|4240118|dbj|BAA74838| (AB007800) NADH-cytochrome b5 reductase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 710: from 1 to 164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 711
- Ceres seq_id 1567741
- Location of start within SEQ ID NO 708: at 523 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase FAD/NAD-binding domain
- Location within SEQ ID NO 711: from 31 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 335
- gi No. 4240116

- Description: (AB007799) NADH-cytochrome b5 reductase [Arabidopsis thaliana] >gi|4240118|dbj|BAA74838| (AB007800) NADH-cytochrome b5 reductase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 711: from 1 to 158

Maximum Length Sequence:
related to:
Clone IDs:
34681
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 712
- Ceres seq_id 1567750

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 713
- Ceres seq_id 1567751
- Location of start within SEQ ID NO 712: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 713: from 42 to 190 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 336
- gi No. 2347208
- Description: (AC002338) APG protein isolog [Arabidopsis thaliana]
- % Identity: 72.8
- Alignment Length: 360
- Location of Alignment in SEQ ID NO 713: from 8 to 367

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 714
- Ceres seq_id 1567752
- Location of start within SEQ ID NO 712: at 24 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 714: from 35 to 183 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 337
- gi No. 2347208
- Description: (AC002338) APG protein isolog [Arabidopsis thaliana]
- % Identity: 72.8
- Alignment Length: 360
- Location of Alignment in SEQ ID NO 714: from 1 to 360

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 715
- Ceres seq_id 1567753
- Location of start within SEQ ID NO 712: at 447 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 338
- gi No. 2347208
- Description: (AC002338) APG protein isolog [Arabidopsis thaliana]

- % Identity: 72.8
- Alignment Length: 360
- Location of Alignment in SEQ ID NO 715: from 1 to 219

Maximum Length Sequence:
related to:
Clone IDs:
2806
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 716
- Ceres seq_id 1567762
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 717
- Ceres seq_id 1567763
- Location of start within SEQ ID NO 716: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 717: from 28 to 124 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 718
- Ceres seq_id 1567764
- Location of start within SEQ ID NO 716: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 718: from 17 to 113 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 719
- Ceres seq_id 1567765
- Location of start within SEQ ID NO 716: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 719: from 3 to 99 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
14136
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 720
- Ceres seq_id 1567773
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 721
- Ceres seq_id 1567774
- Location of start within SEQ ID NO 720: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Matrixin
- Location within SEQ ID NO 721: from 79 to 257 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 339
- gi No. 2245019
- Description: (Z97341) proteinase like protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 721: from 4 to 367

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 722
- Ceres seq_id 1567775
- Location of start within SEQ ID NO 720: at 11 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Matrixin
- Location within SEQ ID NO 722: from 76 to 254 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 340
- gi No. 2245019
- Description: (Z97341) proteinase like protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 722: from 1 to 364

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 723
- Ceres seq_id 1567776
- Location of start within SEQ ID NO 720: at 383 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Matrixin
- Location within SEQ ID NO 723: from 1 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 341
- gi No. 2245019
- Description: (Z97341) proteinase like protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 723: from 1 to 240

Maximum Length Sequence:
related to:
Clone IDs:
26831
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 724
- Ceres seq_id 1567777
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 725
- Ceres seq_id 1567778
- Location of start within SEQ ID NO 724: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione peroxidases
- Location within SEQ ID NO 725: from 81 to 190 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 342
- gi No. 2760606
- Description: (AB001568) phospholipid hydroperoxide glutathione peroxidase-like protein [Arabidopsis thaliana] >gi|3004869 (AF030132) glutathione peroxidase; ATGP1 [Arabidopsis thaliana] >gi|4539451|emb|CAB39931.1| (AL049500) phospholipid hydroperoxide
- % Identity: 99.4
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 725: from 73 to 241

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 726
- Ceres seq_id 1567779
- Location of start within SEQ ID NO 724: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione peroxidases
- Location within SEQ ID NO 726: from 72 to 181 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 343
- gi No. 2760606
- Description: (AB001568) phospholipid hydroperoxide glutathione peroxidase-like protein [Arabidopsis thaliana] >gi|3004869 (AF030132) glutathione peroxidase; ATGP1 [Arabidopsis thaliana] >gi|4539451|emb|CAB39931.1| (AL049500) phospholipid hydroperoxide
- % Identity: 99.4
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 726: from 64 to 232

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 727
- Ceres seq_id 1567780
- Location of start within SEQ ID NO 724: at 217 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione peroxidases
- Location within SEQ ID NO 727: from 9 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 344
- gi No. 2760606
- Description: (AB001568) phospholipid hydroperoxide glutathione peroxidase-like protein [Arabidopsis thaliana] >gi|3004869 (AF030132) glutathione peroxidase; ATGP1 [Arabidopsis thaliana] >gi|4539451|emb|CAB39931.1| (AL049500) phospholipid hydroperoxide
- % Identity: 99.4
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 727: from 1 to 169

Maximum Length Sequence:
related to:

Clone IDs:
7365
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 728
- Ceres seq_id 1567785
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 729
- Ceres seq_id 1567786
- Location of start within SEQ ID NO 728: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 345
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 729: from 123 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 730
- Ceres seq_id 1567787
- Location of start within SEQ ID NO 728: at 340 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 346
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 730: from 58 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 731
- Ceres seq_id 1567788
- Location of start within SEQ ID NO 728: at 373 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 347
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 731: from 47 to 79

Maximum Length Sequence:
related to:
Clone IDs:
114256
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 732
- Ceres seq_id 1567789
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 733
- Ceres seq_id 1567790
- Location of start within SEQ ID NO 732: at 193 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 348
- gi No. 2275204
- Description: (AC002337) DNA binding protein isolog [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 337
- Location of Alignment in SEQ ID NO 733: from 1 to 337

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 734
- Ceres seq_id 1567791
- Location of start within SEQ ID NO 732: at 298 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 349
- gi No. 2275204
- Description: (AC002337) DNA binding protein isolog [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 337
- Location of Alignment in SEQ ID NO 734: from 1 to 302

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 735
- Ceres seq_id 1567792
- Location of start within SEQ ID NO 732: at 685 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 350
- gi No. 2275204
- Description: (AC002337) DNA binding protein isolog [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 337
- Location of Alignment in SEQ ID NO 735: from 1 to 173

Maximum Length Sequence:
related to:
Clone IDs:
12801
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 736
- Ceres seq_id 1567797
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 737
- Ceres seq_id 1567798
- Location of start within SEQ ID NO 736: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 737: from 15 to 163 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 351
- gi No. 2129568
- Description: cytosolic cyclophilin ROC2 - Arabidopsis thaliana
>gi|1305457 (U40400) cytosolic cyclophilin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 153
- Location of Alignment in SEQ ID NO 737: from 11 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 738
- Ceres seq_id 1567799
- Location of start within SEQ ID NO 736: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 738: from 5 to 153 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 352
- gi No. 2129568
- Description: cytosolic cyclophilin ROC2 - Arabidopsis thaliana
>gi|1305457 (U40400) cytosolic cyclophilin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 153
- Location of Alignment in SEQ ID NO 738: from 1 to 153

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 739
- Ceres seq_id 1567800
- Location of start within SEQ ID NO 736: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 739: from 1 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 353
- gi No. 2129568
- Description: cytosolic cyclophilin ROC2 - Arabidopsis thaliana
>gi|1305457 (U40400) cytosolic cyclophilin [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 153
- Location of Alignment in SEQ ID NO 739: from 1 to 139

Maximum Length Sequence:
related to:
Clone IDs:
12817
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 740
- Ceres seq_id 1567812
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 741
- Ceres seq_id 1567813
- Location of start within SEQ ID NO 740: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- 14-3-3 proteins
- Location within SEQ ID NO 741: from 36 to 274 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 354
- gi No. 5802794
- Description: (AF145300) 14-3-3 protein GF14 kappa [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 741: from 30 to 277

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 742
- Ceres seq_id 1567814
- Location of start within SEQ ID NO 740: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 14-3-3 proteins
- Location within SEQ ID NO 742: from 7 to 245 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 355
- gi No. 5802794
- Description: (AF145300) 14-3-3 protein GF14 kappa [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 742: from 1 to 248

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 743
- Ceres seq_id 1567815
- Location of start within SEQ ID NO 740: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 14-3-3 proteins
- Location within SEQ ID NO 743: from 1 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 356
- gi No. 5802794
- Description: (AF145300) 14-3-3 protein GF14 kappa [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 743: from 1 to 236

Maximum Length Sequence:
related to:
Clone IDs:
114733
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 744
- Ceres seq_id 1567835

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 745
- Ceres seq_id 1567836
- Location of start within SEQ ID NO 744: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 745: from 256 to 390 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 357
- gi No. 2281085
- Description: (AC002333) CTR1 protein kinase isolog [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 745: from 250 to 390

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 746
- Ceres seq_id 1567837
- Location of start within SEQ ID NO 744: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 746: from 209 to 343 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 358
- gi No. 2281085
- Description: (AC002333) CTR1 protein kinase isolog [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 746: from 203 to 343

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 747
- Ceres seq_id 1567838
- Location of start within SEQ ID NO 744: at 266 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 747: from 168 to 302 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 359
- gi No. 2281085
- Description: (AC002333) CTR1 protein kinase isolog [Arabidopsis thaliana]
- % Identity: 74.8
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 747: from 162 to 302

Maximum Length Sequence:
related to:
Clone IDs:
114761
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 748
- Ceres seq_id 1567843
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 749
- Ceres seq_id 1567844
- Location of start within SEQ ID NO 748: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldehyde dehydrogenase
- Location within SEQ ID NO 749: from 39 to 490 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 360
- gi No. 4835765
- Description: (AC007202) Is a member of the PF|00171 aldehyde dehydrogenase family. ESTs gb|T21534, gb|N65241 and gb|AA395614 come from this gene. [Arabidopsis thaliana]
- % Identity: 96.7
- Alignment Length: 509
- Location of Alignment in SEQ ID NO 749: from 3 to 495

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 750
- Ceres seq_id 1567845
- Location of start within SEQ ID NO 748: at 167 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldehyde dehydrogenase
- Location within SEQ ID NO 750: from 37 to 488 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 361
- gi No. 4835765
- Description: (AC007202) Is a member of the PF|00171 aldehyde dehydrogenase family. ESTs gb|T21534, gb|N65241 and gb|AA395614 come from this gene. [Arabidopsis thaliana]
- % Identity: 96.7
- Alignment Length: 509
- Location of Alignment in SEQ ID NO 750: from 1 to 493

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 751
- Ceres seq_id 1567846
- Location of start within SEQ ID NO 748: at 320 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldehyde dehydrogenase
- Location within SEQ ID NO 751: from 1 to 437 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 362
- gi No. 4835765
- Description: (AC007202) Is a member of the PF|00171 aldehyde dehydrogenase family. ESTs gb|T21534, gb|N65241 and gb|AA395614 come from this gene. [Arabidopsis thaliana]
- % Identity: 96.7
- Alignment Length: 509
- Location of Alignment in SEQ ID NO 751: from 1 to 442

Maximum Length Sequence:
related to:

Clone IDs:
115358
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 752
- Ceres seq_id 1567877
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 753
- Ceres seq_id 1567878
- Location of start within SEQ ID NO 752: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 753: from 113 to 171 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 754
- Ceres seq_id 1567879
- Location of start within SEQ ID NO 752: at 306 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 754: from 12 to 70 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
115406
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 755
- Ceres seq_id 1567884
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 756
- Ceres seq_id 1567885
- Location of start within SEQ ID NO 755: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein
- Location within SEQ ID NO 756: from 23 to 107 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 757
- Ceres seq_id 1567886
- Location of start within SEQ ID NO 755: at 43 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein
- Location within SEQ ID NO 757: from 9 to 93 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 758
- Ceres seq_id 1567887
- Location of start within SEQ ID NO 755: at 194 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
115583

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 759
- Ceres seq_id 1567896

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 760
- Ceres seq_id 1567897
- Location of start within SEQ ID NO 759: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 760: from 1 to 364 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 363
- gi No. 2864626
- Description: (AL021811) cytochrome p450 - like protein [Arabidopsis thaliana] >gi|3858934|emb|CAA16572.1| (AL021636) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 367
- Location of Alignment in SEQ ID NO 760: from 1 to 367

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 761
- Ceres seq_id 1567898
- Location of start within SEQ ID NO 759: at 13 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 761: from 1 to 360 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 364
- gi No. 2864626
- Description: (AL021811) cytochrome p450 - like protein [Arabidopsis thaliana] >gi|3858934|emb|CAA16572.1| (AL021636) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 367
- Location of Alignment in SEQ ID NO 761: from 1 to 363

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 762
- Ceres seq_id 1567899
- Location of start within SEQ ID NO 759: at 22 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 762: from 1 to 357 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 365
- gi No. 2864626
- Description: (AL021811) cytochrome p450 - like protein [Arabidopsis thaliana] >gi|3858934|emb|CAA16572.1| (AL021636) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 367
- Location of Alignment in SEQ ID NO 762: from 1 to 360

Maximum Length Sequence:
related to:
Clone IDs:
115662
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 763
- Ceres seq_id 1567904
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 764
- Ceres seq_id 1567905
- Location of start within SEQ ID NO 763: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 764: from 32 to 179 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 366
- gi No. 3421082
- Description: (AF043523) 20S proteasome subunit PAD2 [Arabidopsis thaliana]
- % Identity: 96
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 764: from 1 to 248

Maximum Length Sequence:
related to:
Clone IDs:
12972
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 765
- Ceres seq_id 1567929
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 766
- Ceres seq_id 1567930
- Location of start within SEQ ID NO 765: at 172 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TCP-1/cpn60 chaperonin family
- Location within SEQ ID NO 766: from 30 to 520 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 767
- Ceres seq_id 1567931
- Location of start within SEQ ID NO 765: at 232 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TCP-1/cpn60 chaperonin family
- Location within SEQ ID NO 767: from 10 to 500 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 768
- Ceres seq_id 1567932
- Location of start within SEQ ID NO 765: at 253 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TCP-1/cpn60 chaperonin family
- Location within SEQ ID NO 768: from 3 to 493 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
116117
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 769
- Ceres seq_id 1567933

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 770
- Ceres seq_id 1567934
- Location of start within SEQ ID NO 769: at 257 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 770: from 242 to 311 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 367
- gi No. 4006877
- Description: (Z99707) RNA-binding like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 770: from 143 to 359

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 771
- Ceres seq_id 1567935
- Location of start within SEQ ID NO 769: at 323 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 771: from 220 to 289 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 368

- gi No. 4006877
- Description: (Z99707) RNA-binding like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 771: from 121 to 337

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 772
- Ceres seq_id 1567936
- Location of start within SEQ ID NO 769: at 359 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 772: from 208 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 369
- gi No. 4006877
- Description: (Z99707) RNA-binding like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 772: from 109 to 325

Maximum Length Sequence:
related to:
Clone IDs:
116238
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 773
- Ceres seq_id 1567941

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 774
- Ceres seq_id 1567942
- Location of start within SEQ ID NO 773: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 774: from 87 to 210 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 775
- Ceres seq_id 1567943
- Location of start within SEQ ID NO 773: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 775: from 67 to 190 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 776
- Ceres seq_id 1567944
- Location of start within SEQ ID NO 773: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 776: from 48 to 171 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
116257
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 777
- Ceres seq_id 1567945
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 778
- Ceres seq_id 1567946
- Location of start within SEQ ID NO 777: at 633 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 370
- gi No. 3915639
- Description: ALTERNATIVE OXIDASE 1A PRECURSOR
>gi|2506083|dbj|BAA22625| (D89875) alternative oxidase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 215
- Location of Alignment in SEQ ID NO 778: from 1 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 779
- Ceres seq_id 1567947
- Location of start within SEQ ID NO 777: at 636 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 371
- gi No. 3915639
- Description: ALTERNATIVE OXIDASE 1A PRECURSOR
>gi|2506083|dbj|BAA22625| (D89875) alternative oxidase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 215
- Location of Alignment in SEQ ID NO 779: from 1 to 174

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 780
- Ceres seq_id 1567948
- Location of start within SEQ ID NO 777: at 666 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 372
- gi No. 3915639
- Description: ALTERNATIVE OXIDASE 1A PRECURSOR
>gi|2506083|dbj|BAA22625| (D89875) alternative oxidase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 215

- Location of Alignment in SEQ ID NO 780: from 1 to 164

Maximum Length Sequence:
related to:
Clone IDs:
116683
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 781
- Ceres seq_id 1567957
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 782
- Ceres seq_id 1567958
- Location of start within SEQ ID NO 781: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 35
- Location within SEQ ID NO 782: from 1 to 229 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 373
- gi No. 3204134
- Description: (AJ006771) beta-galactosidase [Cicer arietinum]
- % Identity: 75.4
- Alignment Length: 272
- Location of Alignment in SEQ ID NO 782: from 1 to 272

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 783
- Ceres seq_id 1567959
- Location of start within SEQ ID NO 781: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 35
- Location within SEQ ID NO 783: from 1 to 224 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 374
- gi No. 3204134
- Description: (AJ006771) beta-galactosidase [Cicer arietinum]
- % Identity: 75.4
- Alignment Length: 272
- Location of Alignment in SEQ ID NO 783: from 1 to 267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 784
- Ceres seq_id 1567960
- Location of start within SEQ ID NO 781: at 1281 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 784: from 47 to 133 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
116764

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 785
- Ceres seq_id 1567964

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 786
- Ceres seq_id 1567965
- Location of start within SEQ ID NO 785: at 546 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 786: from 183 to 237 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 787
- Ceres seq_id 1567966
- Location of start within SEQ ID NO 785: at 549 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 787: from 182 to 236 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 788
- Ceres seq_id 1567967
- Location of start within SEQ ID NO 785: at 561 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 788: from 178 to 232 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
13054

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 789
- Ceres seq_id 1567975

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 790
- Ceres seq_id 1567976
- Location of start within SEQ ID NO 789: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 790: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 375
- gi No. 115767
- Description: CHLOROPHYLL A-B BINDING PROTEIN OF LHCII TYPE I PRECURSOR (CAB-165/180) (LHCP) >gi|81603|pir||A29280 chlorophyll a/b-binding

protein ab165 - Arabidopsis thaliana >gi|16368|emb|CAA27540| (X03907) chlorophyll a/b binding protein thaliana]
- % Identity: 100
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 790: from 1 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 791
- Ceres seq_id 1567977
- Location of start within SEQ ID NO 789: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 791: from 1 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 376
- gi No. 115767
- Description: CHLOROPHYLL A-B BINDING PROTEIN OF LHCII TYPE I PRECURSOR (CAB-165/180) (LHCP) >gi|81603|pir||A29280 chlorophyll a/b-binding protein ab165 - Arabidopsis thaliana >gi|16368|emb|CAA27540| (X03907) chlorophyll a/b binding protein thaliana]
- % Identity: 100
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 791: from 1 to 99

Maximum Length Sequence:
related to:
Clone IDs:
41809
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 792
- Ceres seq_id 1567993
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 793
- Ceres seq_id 1567994
- Location of start within SEQ ID NO 792: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 793: from 92 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 377
- gi No. 4585979
- Description: (AC005287) Similar to ABC-transporter atp-binding protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 793: from 76 to 227

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 794
- Ceres seq_id 1567995
- Location of start within SEQ ID NO 792: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter

- Location within SEQ ID NO 794: from 48 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 378
- gi No. 4585979
- Description: (AC005287) Similar to ABC-transporter atp-binding protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 794: from 32 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 795
- Ceres seq_id 1567996
- Location of start within SEQ ID NO 792: at 283 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 795: from 1 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 379
- gi No. 4585979
- Description: (AC005287) Similar to ABC-transporter atp-binding protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 795: from 1 to 133

Maximum Length Sequence:
related to:
Clone IDs:
42112
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 796
- Ceres seq_id 1567997
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 797
- Ceres seq_id 1567998
- Location of start within SEQ ID NO 796: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ergosterol biosynthesis ERG4/ERG24 family
- Location within SEQ ID NO 797: from 1 to 376 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 380
- gi No. 1245182
- Description: (U49398) sterol delta-7 reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 797: from 1 to 376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 798
- Ceres seq_id 1567999
- Location of start within SEQ ID NO 796: at 371 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ergosterol biosynthesis ERG4/ERG24 family
- Location within SEQ ID NO 798: from 1 to 253 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 381
- gi No. 1245182
- Description: (U49398) sterol delta-7 reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 798: from 1 to 253

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 799
- Ceres seq_id 1568000
- Location of start within SEQ ID NO 796: at 437 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ergosterol biosynthesis ERG4/ERG24 family
- Location within SEQ ID NO 799: from 1 to 231 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 382
- gi No. 1245182
- Description: (U49398) sterol delta-7 reductase [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 799: from 1 to 231

Maximum Length Sequence:
related to:
Clone IDs:
41984
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 800
- Ceres seq_id 1568007
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 801
- Ceres seq_id 1568008
- Location of start within SEQ ID NO 800: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 801: from 2 to 62 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 383
- gi No. 2252844
- Description: (AF013293) belongs to the cytochrome p450 family [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 801: from 2 to 91

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 802

- Ceres seq_id 1568009
- Location of start within SEQ ID NO 800: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 384
- gi No. 2252844
- Description: (AF013293) belongs to the cytochrome p450 family [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 802: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
36505
6124

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 803
- Ceres seq_id 1568029

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 804
- Ceres seq_id 1568030
- Location of start within SEQ ID NO 803: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 804: from 87 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 385
- gi No. 4587545
- Description: (AC006577) Identical to mRNA gb|X71915 from A.thaliana and contains the Kelch motif. ESTs gb|T04568, gb|T21762, gb|H38255, gb|Z18380, gb|R30291, gb|T21120, gb|R90517, gb|T75605, gb|R30422, gb|T04568, gb|T21762, gb|...
- % Identity: 99.4
- Alignment Length: 168
- Location of Alignment in SEQ ID NO 804: from 1 to 167

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 805
- Ceres seq_id 1568031
- Location of start within SEQ ID NO 803: at 323 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 805: from 7 to 58 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 386
- gi No. 4587545
- Description: (AC006577) Identical to mRNA gb|X71915 from A.thaliana and contains the Kelch motif. ESTs gb|T04568, gb|T21762, gb|H38255, gb|Z18380, gb|R30291, gb|T21120, gb|R90517, gb|T75605, gb|R30422, gb|T04568, gb|T21762, gb|...
- % Identity: 99.4

- Alignment Length: 168
- Location of Alignment in SEQ ID NO 805: from 1 to 87

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 806
- Ceres seq_id 1568032
- Location of start within SEQ ID NO 803: at 877 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
117503

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 807
- Ceres seq_id 1568037

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 808
- Ceres seq_id 1568038
- Location of start within SEQ ID NO 807: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SET domain
- Location within SEQ ID NO 808: from 200 to 313 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 809
- Ceres seq_id 1568039
- Location of start within SEQ ID NO 807: at 463 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SET domain
- Location within SEQ ID NO 809: from 97 to 210 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 810
- Ceres seq_id 1568040
- Location of start within SEQ ID NO 807: at 532 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SET domain
- Location within SEQ ID NO 810: from 74 to 187 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
117626

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 811

- Ceres seq_id 1568045

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 812
- Ceres seq_id 1568046
- Location of start within SEQ ID NO 811: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 387
- gi No. 4886307
- Description: (AJ242588) 1-deoxy-d-xylulose-5-phosphate reductoisomerase [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 812: from 72 to 325

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 813
- Ceres seq_id 1568047
- Location of start within SEQ ID NO 811: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 388
- gi No. 4886307
- Description: (AJ242588) 1-deoxy-d-xylulose-5-phosphate reductoisomerase [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 813: from 71 to 324

Maximum Length Sequence:
related to:
Clone IDs:
117866

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 814
- Ceres seq_id 1568058

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 815
- Ceres seq_id 1568059
- Location of start within SEQ ID NO 814: at 637 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 815: from 1 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 389
- gi No. 1709798
- Description: 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG >gi|1155334 (U43398) POTATP1 [Solanum tuberosum]
- % Identity: 91
- Alignment Length: 410
- Location of Alignment in SEQ ID NO 815: from 1 to 205

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 816

- Ceres seq_id 1568060
- Location of start within SEQ ID NO 814: at 733 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 816: from 1 to 143 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 390
- gi No. 1709798
- Description: 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG >gi|1155334 (U43398) POTATP1 [Solanum tuberosum]
- % Identity: 91
- Alignment Length: 410
- Location of Alignment in SEQ ID NO 816: from 1 to 173

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 817
- Ceres seq_id 1568061
- Location of start within SEQ ID NO 814: at 865 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 817: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 391
- gi No. 1709798
- Description: 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG >gi|1155334 (U43398) POTATP1 [Solanum tuberosum]
- % Identity: 91
- Alignment Length: 410
- Location of Alignment in SEQ ID NO 817: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
118351
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 818
- Ceres seq_id 1568087

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 819
- Ceres seq_id 1568088
- Location of start within SEQ ID NO 818: at 42 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 819: from 15 to 231 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
118355
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 820

- Ceres seq_id 1568089

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 821
- Ceres seq_id 1568090
- Location of start within SEQ ID NO 820: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Acetyltransferase (GNAT) family
- Location within SEQ ID NO 821: from 24 to 153 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 822
- Ceres seq_id 1568091
- Location of start within SEQ ID NO 820: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Acetyltransferase (GNAT) family
- Location within SEQ ID NO 822: from 1 to 117 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 823
- Ceres seq_id 1568092
- Location of start within SEQ ID NO 820: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Acetyltransferase (GNAT) family
- Location within SEQ ID NO 823: from 1 to 102 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
118434
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 824
- Ceres seq_id 1568093

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 825
- Ceres seq_id 1568094
- Location of start within SEQ ID NO 824: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 825: from 4 to 301 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 392
- gi No. 1168811
- Description: CELL DIVISION CONTROL PROTEIN 2 HOMOLOG B >gi|282864|pir||S23096 protein kinase (EC 2.7.1.37) cdc2 homolog B - Arabidopsis thaliana >gi|217851|dbj|BAA01624| (D10851) p32 protein serine/threonine kinase-related protein [Arabidopsis thaliana]

- % Identity: 100
- Alignment Length: 309
- Location of Alignment in SEQ ID NO 825: from 1 to 309

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 826
- Ceres seq_id 1568095
- Location of start within SEQ ID NO 824: at 151 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 826: from 1 to 280 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 393
- gi No. 1168811
- Description: CELL DIVISION CONTROL PROTEIN 2 HOMOLOG B >gi|282864|pir||S23096 protein kinase (EC 2.7.1.37) cdc2 homolog B - Arabidopsis thaliana >gi|217851|dbj|BAA01624| (D10851) p32 protein serine/threonine kinase-related protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 309
- Location of Alignment in SEQ ID NO 826: from 1 to 288

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 827
- Ceres seq_id 1568096
- Location of start within SEQ ID NO 824: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 827: from 1 to 263 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 394
- gi No. 1168811
- Description: CELL DIVISION CONTROL PROTEIN 2 HOMOLOG B >gi|282864|pir||S23096 protein kinase (EC 2.7.1.37) cdc2 homolog B - Arabidopsis thaliana >gi|217851|dbj|BAA01624| (D10851) p32 protein serine/threonine kinase-related protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 309
- Location of Alignment in SEQ ID NO 827: from 1 to 271

Maximum Length Sequence:
related to:
Clone IDs:
118626

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 828
- Ceres seq_id 1568100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 829
- Ceres seq_id 1568101
- Location of start within SEQ ID NO 828: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase

- Location within SEQ ID NO 829: from 41 to 308 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 395
- gi No. 1402920
- Description: (X98321) peroxidase [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 829: from 1 to 308

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 830
- Ceres seq_id 1568102
- Location of start within SEQ ID NO 828: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 830: from 38 to 305 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 396
- gi No. 1402920
- Description: (X98321) peroxidase [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 830: from 1 to 305

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 831
- Ceres seq_id 1568103
- Location of start within SEQ ID NO 828: at 127 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 831: from 21 to 288 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 397
- gi No. 1402920
- Description: (X98321) peroxidase [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 831: from 1 to 288

Maximum Length Sequence:
related to:
Clone IDs:
118636
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 832
- Ceres seq_id 1568104

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 833
- Ceres seq_id 1568105
- Location of start within SEQ ID NO 832: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydroorotase-like

- Location within SEQ ID NO 833: from 88 to 246 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 834
- Ceres seq_id 1568106
- Location of start within SEQ ID NO 832: at 5 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydroorotase-like
- Location within SEQ ID NO 834: from 87 to 245 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 835
- Ceres seq_id 1568107
- Location of start within SEQ ID NO 832: at 757 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydroorotase-like
- Location within SEQ ID NO 835: from 4 to 230 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
119129
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 836
- Ceres seq_id 1568124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 837
- Ceres seq_id 1568125
- Location of start within SEQ ID NO 836: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 837: from 49 to 174 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 398
- gi No. 2980791
- Description: (AL022197) actin depolymerizing factor-like protein [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 837: from 45 to 174

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 838
- Ceres seq_id 1568126
- Location of start within SEQ ID NO 836: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 838: from 12 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 399
- gi No. 2980791
- Description: (AL022197) actin depolymerizing factor-like protein [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 838: from 8 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 839
- Ceres seq_id 1568127
- Location of start within SEQ ID NO 836: at 135 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 839: from 5 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 400
- gi No. 2980791
- Description: (AL022197) actin depolymerizing factor-like protein [Arabidopsis thaliana]
- % Identity: 90
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 839: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
119192
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 840
- Ceres seq_id 1568128

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 841
- Ceres seq_id 1568129
- Location of start within SEQ ID NO 840: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 841: from 274 to 308 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 401
- gi No. 1361991
- Description: homeotic protein ATK1 - Arabidopsis thaliana >gi|984046|emb|CAA57122| (X81354) ATK1 [Arabidopsis thaliana] >gi|984048|emb|CAA57121| (X81353) ATK1 [Arabidopsis thaliana]
- % Identity: 72.3
- Alignment Length: 329
- Location of Alignment in SEQ ID NO 841: from 1 to 329

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 842
- Ceres seq_id 1568130

- Location of start within SEQ ID NO 840: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 842: from 271 to 305 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 402
- gi No. 1361991
- Description: homeotic protein ATK1 - Arabidopsis thaliana
>gi|984046|emb|CAA57122| (X81354) ATK1 [Arabidopsis thaliana]
>gi|984048|emb|CAA57121| (X81353) ATK1 [Arabidopsis thaliana]
- % Identity: 72.3
- Alignment Length: 329
- Location of Alignment in SEQ ID NO 842: from 1 to 326

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 843
- Ceres seq_id 1568131
- Location of start within SEQ ID NO 840: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 843: from 255 to 289 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 403
- gi No. 1361991
- Description: homeotic protein ATK1 - Arabidopsis thaliana
>gi|984046|emb|CAA57122| (X81354) ATK1 [Arabidopsis thaliana]
>gi|984048|emb|CAA57121| (X81353) ATK1 [Arabidopsis thaliana]
- % Identity: 72.3
- Alignment Length: 329
- Location of Alignment in SEQ ID NO 843: from 1 to 310

Maximum Length Sequence:
related to:
Clone IDs:
119889
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 844
- Ceres seq_id 1568152

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 845
- Ceres seq_id 1568153
- Location of start within SEQ ID NO 844: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 845: from 117 to 345 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 846
- Ceres seq_id 1568154
- Location of start within SEQ ID NO 844: at 443 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 846: from 1 to 229 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 847
- Ceres seq_id 1568155
- Location of start within SEQ ID NO 844: at 581 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 847: from 1 to 183 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
119977
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 848
- Ceres seq_id 1568160
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 849
- Ceres seq_id 1568161
- Location of start within SEQ ID NO 848: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 849: from 64 to 236 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 404
- gi No. 5668780
- Description: (AC007894) Strong similarity to gb|AF092432 protein phosphatase type 2C from Lotus japonicus. EST gb|T76026 comes from this gene. [Arabidopsis thaliana]
- % Identity: 73
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 849: from 1 to 280

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 850
- Ceres seq_id 1568162
- Location of start within SEQ ID NO 848: at 204 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 850: from 34 to 206 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 405
- gi No. 5668780

- Description: (AC007894) Strong similarity to gb|AF092432 protein phosphatase type 2C from Lotus japonicus. EST gb|T76026 comes from this gene. [Arabidopsis thaliana]
- % Identity: 73
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 850: from 1 to 250

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 851
- Ceres seq_id 1568163
- Location of start within SEQ ID NO 848: at 252 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 851: from 18 to 190 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 406
- gi No. 5668780
- Description: (AC007894) Strong similarity to gb|AF092432 protein phosphatase type 2C from Lotus japonicus. EST gb|T76026 comes from this gene. [Arabidopsis thaliana]
- % Identity: 73
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 851: from 1 to 234

Maximum Length Sequence:
related to:
Clone IDs:
120133
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 852
- Ceres seq_id 1568176

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 853
- Ceres seq_id 1568177
- Location of start within SEQ ID NO 852: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 853: from 264 to 292 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 407
- gi No. 3287688
- Description: (AC003979) Contains similarity to ycf37 gene product gb|1001425 from Synechocystis sp. genome gb|D63999. ESTs gb|T43026, gb|R64902, gb|Z18169 and gb|N37374 come from this gene. [Arabidopsis thaliana]
- % Identity: 84.9
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 853: from 28 to 322

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 854
- Ceres seq_id 1568178
- Location of start within SEQ ID NO 852: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 854: from 242 to 270 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 408
- gi No. 3287688
- Description: (AC003979) Contains similarity to ycf37 gene product gb|1001425 from Synechocystis sp. genome gb|D63999. ESTs gb|T43026, gb|R64902, gb|Z18169 and gb|N37374 come from this gene. [Arabidopsis thaliana]
- % Identity: 84.9
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 854: from 6 to 300

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 855
- Ceres seq_id 1568179
- Location of start within SEQ ID NO 852: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 855: from 237 to 265 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 409
- gi No. 3287688
- Description: (AC003979) Contains similarity to ycf37 gene product gb|1001425 from Synechocystis sp. genome gb|D63999. ESTs gb|T43026, gb|R64902, gb|Z18169 and gb|N37374 come from this gene. [Arabidopsis thaliana]
- % Identity: 84.9
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 855: from 1 to 295

Maximum Length Sequence:
related to:
Clone IDs:
120149
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 856
- Ceres seq_id 1568180

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 857
- Ceres seq_id 1568181
- Location of start within SEQ ID NO 856: at 530 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- UDP-glucoronosyl and UDP-glucosyl transferases
- Location within SEQ ID NO 857: from 98 to 210 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 410
- gi No. 2244907
- Description: (Z97339) indole-3-acetate beta-glucosyltransferase like protein [Arabidopsis thaliana]
- % Identity: 98.6
- Alignment Length: 293

- Location of Alignment in SEQ ID NO 857: from 4 to 296

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 858
- Ceres seq_id 1568182
- Location of start within SEQ ID NO 856: at 677 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- UDP-glucoronosyl and UDP-glucosyl transferases
- Location within SEQ ID NO 858: from 49 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 411
- gi No. 2244907
- Description: (Z97339) indole-3-acetate beta-glucosyltransferase like protein [Arabidopsis thaliana]
- % Identity: 98.6
- Alignment Length: 293
- Location of Alignment in SEQ ID NO 858: from 1 to 247

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 859
- Ceres seq_id 1568183
- Location of start within SEQ ID NO 856: at 731 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- UDP-glucoronosyl and UDP-glucosyl transferases
- Location within SEQ ID NO 859: from 31 to 143 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 412
- gi No. 2244907
- Description: (Z97339) indole-3-acetate beta-glucosyltransferase like protein [Arabidopsis thaliana]
- % Identity: 98.6
- Alignment Length: 293
- Location of Alignment in SEQ ID NO 859: from 1 to 229

Maximum Length Sequence:
related to:
Clone IDs:
120306
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 860
- Ceres seq_id 1568184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 861
- Ceres seq_id 1568185
- Location of start within SEQ ID NO 860: at 302 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 861: from 108 to 300 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 862

- Ceres seq_id 1568186
- Location of start within SEQ ID NO 860: at 515 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 862: from 37 to 229 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 863
- Ceres seq_id 1568187
- Location of start within SEQ ID NO 860: at 551 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 863: from 25 to 217 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
120478
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 864
- Ceres seq_id 1568188
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 865
- Ceres seq_id 1568189
- Location of start within SEQ ID NO 864: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Monooxygenase
- Location within SEQ ID NO 865: from 166 to 377 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 413
- gi No. 3426064
- Description: (AJ007588) monooxygenase [Arabidopsis thaliana] >gi|4467141|emb|CAB37510| (AL035540) monooxygenase 2 (MO2) [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 865: from 18 to 409

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 866
- Ceres seq_id 1568190
- Location of start within SEQ ID NO 864: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Monooxygenase
- Location within SEQ ID NO 866: from 149 to 360 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 414

- gi No. 3426064
- Description: (AJ007588) monooxygenase [Arabidopsis thaliana] >gi|4467141|emb|CAB37510| (AL035540) monooxygenase 2 (MO2) [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 866: from 1 to 392

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 867
- Ceres seq_id 1568191
- Location of start within SEQ ID NO 864: at 224 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Monooxygenase
- Location within SEQ ID NO 867: from 92 to 303 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 415
- gi No. 3426064
- Description: (AJ007588) monooxygenase [Arabidopsis thaliana] >gi|4467141|emb|CAB37510| (AL035540) monooxygenase 2 (MO2) [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 867: from 1 to 335

Maximum Length Sequence:
related to:
Clone IDs:
120488
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 868
- Ceres seq_id 1568192

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 869
- Ceres seq_id 1568193
- Location of start within SEQ ID NO 868: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 416
- gi No. 1360084
- Description: (X97945) Zn finger protein [Nicotiana tabacum]
- % Identity: 70.9
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 869: from 12 to 112

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 870
- Ceres seq_id 1568194
- Location of start within SEQ ID NO 868: at 112 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 417
- gi No. 1360084
- Description: (X97945) Zn finger protein [Nicotiana tabacum]

- % Identity: 70.9
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 870: from 1 to 98

Maximum Length Sequence:
related to:
Clone IDs:
120514
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 871
- Ceres seq_id 1568199

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 872
- Ceres seq_id 1568200
- Location of start within SEQ ID NO 871: at 555 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 872: from 1 to 178 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 418
- gi No. 2351580
- Description: (U82433) thymidine diphospho-glucose 4-6-dehydratase homolog [Prunus armeniaca]
- % Identity: 93.8
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 872: from 1 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 873
- Ceres seq_id 1568201
- Location of start within SEQ ID NO 871: at 723 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 873: from 1 to 122 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 419
- gi No. 2351580
- Description: (U82433) thymidine diphospho-glucose 4-6-dehydratase homolog [Prunus armeniaca]
- % Identity: 93.8
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 873: from 1 to 122

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 874
- Ceres seq_id 1568202
- Location of start within SEQ ID NO 871: at 795 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 874: from 1 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 420

- gi No. 2351580
- Description: (U82433) thymidine diphospho-glucose 4-6-dehydratase homolog [Prunus armeniaca]
- % Identity: 93.8
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 874: from 1 to 98

Maximum Length Sequence:
related to:
Clone IDs:
120919
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 875
- Ceres seq_id 1568213
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 876
- Ceres seq_id 1568214
- Location of start within SEQ ID NO 875: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 876: from 188 to 308 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 877
- Ceres seq_id 1568215
- Location of start within SEQ ID NO 875: at 1029 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 877: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 421
- gi No. 98948
- Description: tylosin-resistance protein - Streptomyces fradiae (fragments)
- % Identity: 70
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 877: from 10 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 878
- Ceres seq_id 1568216
- Location of start within SEQ ID NO 875: at 1071 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 878: from 1 to 67 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 422
- gi No. 98948
- Description: tylosin-resistance protein - Streptomyces fradiae (fragments)
- % Identity: 70

- Alignment Length: 31
- Location of Alignment in SEQ ID NO 878: from 1 to 25

Maximum Length Sequence:
related to:
Clone IDs:
120947
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 879
- Ceres seq_id 1568217
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 880
- Ceres seq_id 1568218
- Location of start within SEQ ID NO 879: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 423
- gi No. 120943
- Description: GLUTAMIC ACID-RICH PROTEIN PRECURSOR >gi|627054|pir||A54514 glutamic acid-rich protein precursor - Plasmodium falciparum >gi|160299 (J03998) glutamic acid-rich protein [Plasmodium falciparum]
- % Identity: 78.6
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 880: from 100 to 127

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 881
- Ceres seq_id 1568219
- Location of start within SEQ ID NO 879: at 16 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 424
- gi No. 120943
- Description: GLUTAMIC ACID-RICH PROTEIN PRECURSOR >gi|627054|pir||A54514 glutamic acid-rich protein precursor - Plasmodium falciparum >gi|160299 (J03998) glutamic acid-rich protein [Plasmodium falciparum]
- % Identity: 78.6
- Alignment Length: 28
- Location of Alignment in SEQ ID NO 881: from 95 to 122

Maximum Length Sequence:
related to:
Clone IDs:
121286
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 882
- Ceres seq_id 1568230
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 883
- Ceres seq_id 1568231
- Location of start within SEQ ID NO 882: at 276 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 425
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 883: from 75 to 355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 884
- Ceres seq_id 1568232
- Location of start within SEQ ID NO 882: at 318 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 426
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 884: from 61 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 885
- Ceres seq_id 1568233
- Location of start within SEQ ID NO 882: at 384 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 427
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.6
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 885: from 39 to 319

Maximum Length Sequence:
related to:
Clone IDs:
13503
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 886
- Ceres seq_id 1568234

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 887
- Ceres seq_id 1568235
- Location of start within SEQ ID NO 886: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 887: from 23 to 102 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 888
- Ceres seq_id 1568236
- Location of start within SEQ ID NO 886: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 888: from 8 to 87 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 889
- Ceres seq_id 1568237
- Location of start within SEQ ID NO 886: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 889: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
121993
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 890
- Ceres seq_id 1568263
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 891
- Ceres seq_id 1568264
- Location of start within SEQ ID NO 890: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 428
- gi No. 4884938
- Description: (AF141662) AtHVA22d [Arabidopsis thaliana] >gi|4884948|gb|AAD31887.1|AF141979_1 (AF141979) AtHVA22d [Arabidopsis thaliana]
- % Identity: 71.6
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 891: from 17 to 132

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 892
- Ceres seq_id 1568265
- Location of start within SEQ ID NO 890: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 429
- gi No. 4884938
- Description: (AF141662) AtHVA22d [Arabidopsis thaliana] >gi|4884948|gb|AAD31887.1|AF141979_1 (AF141979) AtHVA22d [Arabidopsis thaliana]
- % Identity: 71.6
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 892: from 1 to 116

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 893
- Ceres seq_id 1568266
- Location of start within SEQ ID NO 890: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 430
- gi No. 4884938
- Description: (AF141662) AtHVA22d [Arabidopsis thaliana] >gi|4884948|gb|AAD31887.1|AF141979_1 (AF141979) AtHVA22d [Arabidopsis thaliana]
- % Identity: 71.6
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 893: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
122177

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 894
- Ceres seq_id 1568267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 895
- Ceres seq_id 1568268
- Location of start within SEQ ID NO 894: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphoglucomutase/phosphomannomutase
- Location within SEQ ID NO 895: from 37 to 417 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 431
- gi No. 2829893
- Description: (AC002311) phosphoglucomutase [Arabidopsis thaliana]
- % Identity: 90.7
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 895: from 22 to 417

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 896
- Ceres seq_id 1568269
- Location of start within SEQ ID NO 894: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphoglucomutase/phosphomannomutase
- Location within SEQ ID NO 896: from 19 to 399 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 432
- gi No. 2829893
- Description: (AC002311) phosphoglucomutase [Arabidopsis thaliana]
- % Identity: 90.7
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 896: from 4 to 399

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 897

- Ceres seq_id 1568270
- Location of start within SEQ ID NO 894: at 288 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphoglucomutase/phosphomannomutase
- Location within SEQ ID NO 897: from 1 to 322 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 433
- gi No. 2829893
- Description: (AC002311) phosphoglucomutase [Arabidopsis thaliana]
- % Identity: 90.7
- Alignment Length: 582
- Location of Alignment in SEQ ID NO 897: from 1 to 322

Maximum Length Sequence:
related to:
Clone IDs:
122537
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 898
- Ceres seq_id 1568286

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 899
- Ceres seq_id 1568287
- Location of start within SEQ ID NO 898: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 899: from 102 to 334 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 434
- gi No. 545577
- Description: tryptophanyl-tRNA-synthetase {EC 6.1.1.2} [cattle, Peptide Partial, 20 aa, segment 4 of 5]
- % Identity: 73.7
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 899: from 255 to 273

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 900
- Ceres seq_id 1568288
- Location of start within SEQ ID NO 898: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 900: from 78 to 310 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 435
- gi No. 545577
- Description: tryptophanyl-tRNA-synthetase {EC 6.1.1.2} [cattle, Peptide Partial, 20 aa, segment 4 of 5]
- % Identity: 73.7
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 900: from 231 to 249

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 901
- Ceres seq_id 1568289
- Location of start within SEQ ID NO 898: at 403 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class I (W and Y)
- Location within SEQ ID NO 901: from 1 to 200 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 436
- gi No. 545577
- Description: tryptophanyl-tRNA-synthetase {EC 6.1.1.2} [cattle, Peptide Partial, 20 aa, segment 4 of 5]
- % Identity: 73.7
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 901: from 121 to 139

Maximum Length Sequence:
related to:
Clone IDs:
13667
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 902
- Ceres seq_id 1568300
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 903
- Ceres seq_id 1568301
- Location of start within SEQ ID NO 902: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adenylate kinase
- Location within SEQ ID NO 903: from 38 to 224 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 437
- gi No. 3746809
- Description: (AF082882) adenylate kinase [Arabidopsis thaliana]
- % Identity: 96.3
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 903: from 1 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 904
- Ceres seq_id 1568302
- Location of start within SEQ ID NO 902: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adenylate kinase
- Location within SEQ ID NO 904: from 20 to 206 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 438
- gi No. 3746809
- Description: (AF082882) adenylate kinase [Arabidopsis thaliana]
- % Identity: 96.3
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 904: from 1 to 228

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 905
- Ceres seq_id 1568303
- Location of start within SEQ ID NO 902: at 246 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adenylate kinase
- Location within SEQ ID NO 905: from 1 to 158 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 439
- gi No. 3746809
- Description: (AF082882) adenylate kinase [Arabidopsis thaliana]
- % Identity: 96.3
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 905: from 1 to 180

Maximum Length Sequence:
related to:
Clone IDs:
123030
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 906
- Ceres seq_id 1568304
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 907
- Ceres seq_id 1568305
- Location of start within SEQ ID NO 906: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 440
- gi No. 2809244
- Description: (AC002560) F21B7.13 [Arabidopsis thaliana]
- % Identity: 92.2
- Alignment Length: 313
- Location of Alignment in SEQ ID NO 907: from 1 to 308

Maximum Length Sequence:
related to:
Clone IDs:
123078
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 908
- Ceres seq_id 1568309
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 909
- Ceres seq_id 1568310
- Location of start within SEQ ID NO 908: at 448 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bromodomain
- Location within SEQ ID NO 909: from 152 to 200 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 910
- Ceres seq_id 1568311
- Location of start within SEQ ID NO 908: at 487 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bromodomain
- Location within SEQ ID NO 910: from 139 to 187 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 911
- Ceres seq_id 1568312
- Location of start within SEQ ID NO 908: at 553 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bromodomain
- Location within SEQ ID NO 911: from 117 to 165 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
13684
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 912
- Ceres seq_id 1568313

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 913
- Ceres seq_id 1568314
- Location of start within SEQ ID NO 912: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 441
- gi No. 3243274
- Description: (AF072134) TCP3 [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 87
- Location of Alignment in SEQ ID NO 913: from 2 to 82

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 914
- Ceres seq_id 1568315
- Location of start within SEQ ID NO 912: at 11 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 442
- gi No. 3243274
- Description: (AF072134) TCP3 [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 87
- Location of Alignment in SEQ ID NO 914: from 1 to 79

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 915
- Ceres seq_id 1568316
- Location of start within SEQ ID NO 912: at 504 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
123200
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 916
- Ceres seq_id 1568321

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 917
- Ceres seq_id 1568322
- Location of start within SEQ ID NO 916: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 443
- gi No. 1483218
- Description: (X99793) induced upon wounding stress [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 41
- Location of Alignment in SEQ ID NO 917: from 2 to 42

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 918
- Ceres seq_id 1568323
- Location of start within SEQ ID NO 916: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 444
- gi No. 1483218
- Description: (X99793) induced upon wounding stress [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 41
- Location of Alignment in SEQ ID NO 918: from 1 to 37

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 919
- Ceres seq_id 1568324
- Location of start within SEQ ID NO 916: at 20 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 445
- gi No. 1483218
- Description: (X99793) induced upon wounding stress [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 41

- Location of Alignment in SEQ ID NO 919: from 1 to 36

Maximum Length Sequence:
related to:
Clone IDs:
13704
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 920
- Ceres seq_id 1568333
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 921
- Ceres seq_id 1568334
- Location of start within SEQ ID NO 920: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 446
- gi No. 3372233
- Description: (AF019248) RNA polymerase I, II and III 24.3 kDa subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 921: from 39 to 157

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 922
- Ceres seq_id 1568335
- Location of start within SEQ ID NO 920: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 447
- gi No. 3372233
- Description: (AF019248) RNA polymerase I, II and III 24.3 kDa subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 922: from 1 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 923
- Ceres seq_id 1568336
- Location of start within SEQ ID NO 920: at 167 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 448
- gi No. 3372233
- Description: (AF019248) RNA polymerase I, II and III 24.3 kDa subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 923: from 1 to 102

Maximum Length Sequence:
related to:
Clone IDs:
123997

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 924
- Ceres seq_id 1568357

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 925
- Ceres seq_id 1568358
- Location of start within SEQ ID NO 924: at 203 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 925: from 122 to 170 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 926
- Ceres seq_id 1568359
- Location of start within SEQ ID NO 924: at 344 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 926: from 75 to 123 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 927
- Ceres seq_id 1568360
- Location of start within SEQ ID NO 924: at 470 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 927: from 33 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
124067

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 928
- Ceres seq_id 1568365

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 929
- Ceres seq_id 1568366
- Location of start within SEQ ID NO 928: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sulfotransferase proteins
- Location within SEQ ID NO 929: from 68 to 329 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 930
- Ceres seq_id 1568367

- Location of start within SEQ ID NO 928: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sulfotransferase proteins
- Location within SEQ ID NO 930: from 63 to 324 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 931
- Ceres seq_id 1568368
- Location of start within SEQ ID NO 928: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sulfotransferase proteins
- Location within SEQ ID NO 931: from 58 to 319 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
124299
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 932
- Ceres seq_id 1568376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 933
- Ceres seq_id 1568377
- Location of start within SEQ ID NO 932: at 176 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 933: from 12 to 72 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 449
- gi No. 1408473
- Description: (U48939) actin depolymerizing factor 2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 933: from 1 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 934
- Ceres seq_id 1568378
- Location of start within SEQ ID NO 932: at 197 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 934: from 5 to 65 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 450
- gi No. 1408473

- Description: (U48939) actin depolymerizing factor 2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 934: from 1 to 65

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 935
- Ceres seq_id 1568379
- Location of start within SEQ ID NO 932: at 327 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 935: from 23 to 87 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 451
- gi No. 1408473
- Description: (U48939) actin depolymerizing factor 2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 935: from 23 to 87

Maximum Length Sequence:
related to:
Clone IDs:
125010
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 936
- Ceres seq_id 1568405

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 937
- Ceres seq_id 1568406
- Location of start within SEQ ID NO 936: at 243 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 938
- Ceres seq_id 1568407
- Location of start within SEQ ID NO 936: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 939
- Ceres seq_id 1568408
- Location of start within SEQ ID NO 936: at 776 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GATA zinc finger
- Location within SEQ ID NO 939: from 41 to 79 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
13876
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 940
- Ceres seq_id 1568409
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 941
- Ceres seq_id 1568410
- Location of start within SEQ ID NO 940: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 941: from 28 to 175 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 942
- Ceres seq_id 1568411
- Location of start within SEQ ID NO 940: at 216 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 942: from 1 to 127 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 943
- Ceres seq_id 1568412
- Location of start within SEQ ID NO 940: at 255 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 943: from 1 to 114 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22292
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 944
- Ceres seq_id 1568425
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 945
- Ceres seq_id 1568426
- Location of start within SEQ ID NO 944: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 945: from 12 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 452
- gi No. 3157925
- Description: (AC002131) ESTs gb|F13979, gb|AA042648, gb|H77240, gb|T76189, gb|H35952, gb|T22022 and gb|T42279 come from this gene. A similar ORF is present on the opposite strand. [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 945: from 12 to 146

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 946
- Ceres seq_id 1568427
- Location of start within SEQ ID NO 944: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 946: from 5 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 453
- gi No. 3157925
- Description: (AC002131) ESTs gb|F13979, gb|AA042648, gb|H77240, gb|T76189, gb|H35952, gb|T22022 and gb|T42279 come from this gene. A similar ORF is present on the opposite strand. [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 946: from 5 to 139

Maximum Length Sequence:
related to:
Clone IDs:
19247
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 947
- Ceres seq_id 1568436

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 948
- Ceres seq_id 1568437
- Location of start within SEQ ID NO 947: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 454
- gi No. 3335335
- Description: (AC004512) ESTs gb|F14113 and gb|T42122 come from this region. [Arabidopsis thaliana]
- % Identity: 97
- Alignment Length: 236
- Location of Alignment in SEQ ID NO 948: from 26 to 256

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 949
- Ceres seq_id 1568438
- Location of start within SEQ ID NO 947: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 455
- gi No. 3335335
- Description: (AC004512) ESTs gb|F14113 and gb|T42122 come from this region. [Arabidopsis thaliana]
- % Identity: 97
- Alignment Length: 236
- Location of Alignment in SEQ ID NO 949: from 1 to 231

Maximum Length Sequence:
related to:
Clone IDs:
125679
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 950
- Ceres seq_id 1568490
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 951
- Ceres seq_id 1568491
- Location of start within SEQ ID NO 950: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 456
- gi No. 1946365
- Description: (U93215) glutaredoxin isolog [Arabidopsis thaliana]
- % Identity: 86.6
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 951: from 12 to 108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 952
- Ceres seq_id 1568492
- Location of start within SEQ ID NO 950: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 457
- gi No. 1946365
- Description: (U93215) glutaredoxin isolog [Arabidopsis thaliana]
- % Identity: 86.6
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 952: from 1 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 953
- Ceres seq_id 1568493
- Location of start within SEQ ID NO 950: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 458
- gi No. 1946365
- Description: (U93215) glutaredoxin isolog [Arabidopsis thaliana]
- % Identity: 86.6
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 953: from 1 to 93

Maximum Length Sequence:

related to:
Clone IDs:
125951
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 954
- Ceres seq_id 1568494
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 955
- Ceres seq_id 1568495
- Location of start within SEQ ID NO 954: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 956
- Ceres seq_id 1568496
- Location of start within SEQ ID NO 954: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 956: from 36 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 459
- gi No. 4768968
- Description: (AF140219) nuclear cap-binding protein; CBP20 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 956: from 1 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 957
- Ceres seq_id 1568497
- Location of start within SEQ ID NO 954: at 646 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
126035
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 958
- Ceres seq_id 1568508
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 959
- Ceres seq_id 1568509
- Location of start within SEQ ID NO 958: at 191 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonucleotide reductases
- Location within SEQ ID NO 959: from 21 to 209 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 460
- gi No. 1710401
- Description: RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (RIBONUCLEOTIDE REDUCTASE) (R2 SUBUNIT) >gi|1044912|emb|CAA63194| (X92443) ribonucleotide reductase R2 [Nicotiana tabacum]
- % Identity: 83.9
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 959: from 1 to 209

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 960
- Ceres seq_id 1568510
- Location of start within SEQ ID NO 958: at 224 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonucleotide reductases
- Location within SEQ ID NO 960: from 10 to 198 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 461
- gi No. 1710401
- Description: RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (RIBONUCLEOTIDE REDUCTASE) (R2 SUBUNIT) >gi|1044912|emb|CAA63194| (X92443) ribonucleotide reductase R2 [Nicotiana tabacum]
- % Identity: 83.9
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 960: from 1 to 198

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 961
- Ceres seq_id 1568511
- Location of start within SEQ ID NO 958: at 398 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonucleotide reductases
- Location within SEQ ID NO 961: from 1 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 462
- gi No. 1710401
- Description: RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (RIBONUCLEOTIDE REDUCTASE) (R2 SUBUNIT) >gi|1044912|emb|CAA63194| (X92443) ribonucleotide reductase R2 [Nicotiana tabacum]
- % Identity: 83.9
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 961: from 1 to 140

Maximum Length Sequence:
related to:
Clone IDs:
126070
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 962
- Ceres seq_id 1568512

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 963
- Ceres seq_id 1568513
- Location of start within SEQ ID NO 962: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 963: from 3 to 288 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 463
- gi No. 5734785
- Description: (AC007980) phosphoprotein phosphatase [Arabidopsis thaliana]
- % Identity: 98
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 963: from 1 to 303

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 964
- Ceres seq_id 1568514
- Location of start within SEQ ID NO 962: at 306 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 964: from 1 to 230 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 464
- gi No. 5734785
- Description: (AC007980) phosphoprotein phosphatase [Arabidopsis thaliana]
- % Identity: 98
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 964: from 1 to 245

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 965
- Ceres seq_id 1568515
- Location of start within SEQ ID NO 962: at 357 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 965: from 1 to 213 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 465
- gi No. 5734785
- Description: (AC007980) phosphoprotein phosphatase [Arabidopsis thaliana]
- % Identity: 98
- Alignment Length: 303
- Location of Alignment in SEQ ID NO 965: from 1 to 228

Maximum Length Sequence:
related to:
Clone IDs:
126460
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 966
- Ceres seq_id 1568528
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 967

- Ceres seq_id 1568529
- Location of start within SEQ ID NO 966: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 968
- Ceres seq_id 1568530
- Location of start within SEQ ID NO 966: at 224 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 968: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 969
- Ceres seq_id 1568531
- Location of start within SEQ ID NO 966: at 308 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 969: from 1 to 71 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
14241
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 970
- Ceres seq_id 1568576

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 971
- Ceres seq_id 1568577
- Location of start within SEQ ID NO 970: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 466
- gi No. 99847
- Description: serine proteinase inhibitor - white mustard >gi|260921|bbs|119279 mustard trypsin inhibitor 2, MTI-2=serine proteinase inhibitor [Sinapis alba=white mustard, seed, Peptide, 63 aa]
- % Identity: 73
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 971: from 30 to 92

Maximum Length Sequence:
related to:
Clone IDs:
14401
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 972

- Ceres seq_id 1568583

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 973
- Ceres seq_id 1568584
- Location of start within SEQ ID NO 972: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 973: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 467
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 92
- Alignment Length: 114
- Location of Alignment in SEQ ID NO 973: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 974
- Ceres seq_id 1568585
- Location of start within SEQ ID NO 972: at 158 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 974: from 1 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 468
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 92
- Alignment Length: 114
- Location of Alignment in SEQ ID NO 974: from 1 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 975
- Ceres seq_id 1568586
- Location of start within SEQ ID NO 972: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 975: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 469
- gi No. 2252857
- Description: (AF013294) similar to acidic ribosomal protein p1 [Arabidopsis thaliana]
- % Identity: 92
- Alignment Length: 114
- Location of Alignment in SEQ ID NO 975: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:

14503
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 976
- Ceres seq_id 1568595
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 977
- Ceres seq_id 1568596
- Location of start within SEQ ID NO 976: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/manganese superoxide dismutases (SODM)
- Location within SEQ ID NO 977: from 30 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 470
- gi No. 3273751
- Description: (AF061518) manganese superoxide dismutase [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 231
- Location of Alignment in SEQ ID NO 977: from 1 to 231

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 978
- Ceres seq_id 1568597
- Location of start within SEQ ID NO 976: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/manganese superoxide dismutases (SODM)
- Location within SEQ ID NO 978: from 1 to 177 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 471
- gi No. 3273751
- Description: (AF061518) manganese superoxide dismutase [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 231
- Location of Alignment in SEQ ID NO 978: from 1 to 180

Maximum Length Sequence:
related to:
Clone IDs:
14562
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 979
- Ceres seq_id 1568598
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 980
- Ceres seq_id 1568599
- Location of start within SEQ ID NO 979: at 274 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SNAP-25 family
- Location within SEQ ID NO 980: from 69 to 252 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 472

- gi No. 5731763
- Description: (X92419) SNAP25A protein [Arabidopsis thaliana] >gi|5731764|emb|CAB52583.1| (X92420) SNAP25AB protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 980: from 10 to 271

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 981
- Ceres seq_id 1568600
- Location of start within SEQ ID NO 979: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SNAP-25 family
- Location within SEQ ID NO 981: from 68 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 473
- gi No. 5731763
- Description: (X92419) SNAP25A protein [Arabidopsis thaliana] >gi|5731764|emb|CAB52583.1| (X92420) SNAP25AB protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 981: from 9 to 270

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 982
- Ceres seq_id 1568601
- Location of start within SEQ ID NO 979: at 346 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SNAP-25 family
- Location within SEQ ID NO 982: from 45 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 474
- gi No. 5731763
- Description: (X92419) SNAP25A protein [Arabidopsis thaliana] >gi|5731764|emb|CAB52583.1| (X92420) SNAP25AB protein [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 982: from 1 to 247

Maximum Length Sequence:
related to:
Clone IDs:
14564
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 983
- Ceres seq_id 1568602

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 984
- Ceres seq_id 1568603
- Location of start within SEQ ID NO 983: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L24e
- Location within SEQ ID NO 984: from 3 to 73 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 475
- gi No. 2961300
- Description: (AJ225027) ribosomal protein L24 [Cicer arietinum]
- % Identity: 88.3
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 984: from 1 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 985
- Ceres seq_id 1568604
- Location of start within SEQ ID NO 983: at 212 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 476
- gi No. 2961300
- Description: (AJ225027) ribosomal protein L24 [Cicer arietinum]
- % Identity: 88.3
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 985: from 1 to 108

Maximum Length Sequence:
related to:
Clone IDs:
14650
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 986
- Ceres seq_id 1568605
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 987
- Ceres seq_id 1568606
- Location of start within SEQ ID NO 986: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 988
- Ceres seq_id 1568607
- Location of start within SEQ ID NO 986: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 477
- gi No. 1769905
- Description: (X98108) 23 kDa polypeptide of oxygen-evolving comlex (OEC) [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 43
- Location of Alignment in SEQ ID NO 988: from 1 to 43

Maximum Length Sequence:
related to:
Clone IDs:
14874
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 989
- Ceres seq_id 1568614

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 990
- Ceres seq_id 1568615
- Location of start within SEQ ID NO 989: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 478
- gi No. 2736147
- Description: (AF021804) fatty acid hydroxylase Fah1p [Arabidopsis thaliana] >gi|3132481 (AC003096) fatty acid hydroxylase, FAH1 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 237
- Location of Alignment in SEQ ID NO 990: from 1 to 236

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 991
- Ceres seq_id 1568616
- Location of start within SEQ ID NO 989: at 342 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 479
- gi No. 2736147
- Description: (AF021804) fatty acid hydroxylase Fah1p [Arabidopsis thaliana] >gi|3132481 (AC003096) fatty acid hydroxylase, FAH1 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 237
- Location of Alignment in SEQ ID NO 991: from 1 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 992
- Ceres seq_id 1568617
- Location of start within SEQ ID NO 989: at 381 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 480
- gi No. 2736147
- Description: (AF021804) fatty acid hydroxylase Fah1p [Arabidopsis thaliana] >gi|3132481 (AC003096) fatty acid hydroxylase, FAH1 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 237
- Location of Alignment in SEQ ID NO 992: from 1 to 148

Maximum Length Sequence:
related to:
Clone IDs:
15025

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 993
- Ceres seq_id 1568635

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 994
- Ceres seq_id 1568636
- Location of start within SEQ ID NO 993: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 994: from 41 to 224 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 995
- Ceres seq_id 1568637
- Location of start within SEQ ID NO 993: at 9 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 995: from 39 to 222 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 996
- Ceres seq_id 1568638
- Location of start within SEQ ID NO 993: at 420 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 996: from 1 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
36621
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 997
- Ceres seq_id 1568645

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 998
- Ceres seq_id 1568646
- Location of start within SEQ ID NO 997: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 998: from 7 to 291 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 481
- gi No. 585629
- Description: SERINE/THREONINE PROTEIN PHOSPHATASE PP2A-2 CATALYTIC SUBUNIT >gi|418778|pir||S31161 phosphoprotein phosphatase (EC 3.1.3.16) 2A-alpha catalytic chain (clone EP8a) - Arabidopsis thaliana >gi|166821 (M96732) protein phosphatase [Arabidopsis
- % Identity: 99.7
- Alignment Length: 306

- Location of Alignment in SEQ ID NO 998: from 1 to 306

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 999
- Ceres seq_id 1568647
- Location of start within SEQ ID NO 997: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 999: from 1 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 482
- gi No. 585629
- Description: SERINE/THREONINE PROTEIN PHOSPHATASE PP2A-2 CATALYTIC SUBUNIT >gi|418778|pir||S31161 phosphoprotein phosphatase (EC 3.1.3.16) 2A-alpha catalytic chain (clone EP8a) - Arabidopsis thaliana >gi|166821 (M96732) protein phosphatase [Arabidopsis
- % Identity: 99.7
- Alignment Length: 306
- Location of Alignment in SEQ ID NO 999: from 1 to 292

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1000
- Ceres seq_id 1568648
- Location of start within SEQ ID NO 997: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ser/Thr protein phosphatase
- Location within SEQ ID NO 1000: from 1 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 483
- gi No. 585629
- Description: SERINE/THREONINE PROTEIN PHOSPHATASE PP2A-2 CATALYTIC SUBUNIT >gi|418778|pir||S31161 phosphoprotein phosphatase (EC 3.1.3.16) 2A-alpha catalytic chain (clone EP8a) - Arabidopsis thaliana >gi|166821 (M96732) protein phosphatase [Arabidopsis
- % Identity: 99.7
- Alignment Length: 306
- Location of Alignment in SEQ ID NO 1000: from 1 to 227

Maximum Length Sequence:
related to:
Clone IDs:
40273
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1001
- Ceres seq_id 1568657

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1002
- Ceres seq_id 1568658
- Location of start within SEQ ID NO 1001: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 1002: from 33 to 310 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 484
- gi No. 5059115
- Description: (AF132212) OPDA-reductase homolog [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 391
- Location of Alignment in SEQ ID NO 1002: from 22 to 412

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1003
- Ceres seq_id 1568659
- Location of start within SEQ ID NO 1001: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 1003: from 12 to 289 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 485
- gi No. 5059115
- Description: (AF132212) OPDA-reductase homolog [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 391
- Location of Alignment in SEQ ID NO 1003: from 1 to 391

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1004
- Ceres seq_id 1568660
- Location of start within SEQ ID NO 1001: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 1004: from 1 to 272 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 486
- gi No. 5059115
- Description: (AF132212) OPDA-reductase homolog [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 391
- Location of Alignment in SEQ ID NO 1004: from 1 to 374

Maximum Length Sequence:
related to:
Clone IDs:
39195
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1005
- Ceres seq_id 1568669

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1006
- Ceres seq_id 1568670
- Location of start within SEQ ID NO 1005: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Peroxidase
- Location within SEQ ID NO 1006: from 56 to 333 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 487
- gi No. 1402900
- Description: (X98322) peroxidase [Arabidopsis thaliana]
>gi|1429219|emb|CAA67312| (X98776) peroxidase ATP13a [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 313
- Location of Alignment in SEQ ID NO 1006: from 21 to 333

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1007
- Ceres seq_id 1568671
- Location of start within SEQ ID NO 1005: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1007: from 36 to 313 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 488
- gi No. 1402900
- Description: (X98322) peroxidase [Arabidopsis thaliana]
>gi|1429219|emb|CAA67312| (X98776) peroxidase ATP13a [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 313
- Location of Alignment in SEQ ID NO 1007: from 1 to 313

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1008
- Ceres seq_id 1568672
- Location of start within SEQ ID NO 1005: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1008: from 35 to 312 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 489
- gi No. 1402900
- Description: (X98322) peroxidase [Arabidopsis thaliana]
>gi|1429219|emb|CAA67312| (X98776) peroxidase ATP13a [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 313
- Location of Alignment in SEQ ID NO 1008: from 1 to 312

Maximum Length Sequence:
related to:
Clone IDs:
24151
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1009
- Ceres seq_id 1568673
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1010
- Ceres seq_id 1568674
- Location of start within SEQ ID NO 1009: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 1010: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 490
- gi No. 2104691
- Description: (U92794) alpha glucosidase II, beta subunit [Mus musculus]
- % Identity: 76.2
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1010: from 100 to 120

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1011
- Ceres seq_id 1568675
- Location of start within SEQ ID NO 1009: at 256 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 491
- gi No. 2104691
- Description: (U92794) alpha glucosidase II, beta subunit [Mus musculus]
- % Identity: 76.2
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1011: from 35 to 55

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1012
- Ceres seq_id 1568676
- Location of start within SEQ ID NO 1009: at 280 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 492
- gi No. 2104691
- Description: (U92794) alpha glucosidase II, beta subunit [Mus musculus]
- % Identity: 76.2
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1012: from 27 to 47

Maximum Length Sequence:
related to:
Clone IDs:
141687

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1013
- Ceres seq_id 1568683

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1014
- Ceres seq_id 1568684
- Location of start within SEQ ID NO 1013: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Ank repeat
- Location within SEQ ID NO 1014: from 166 to 199 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1015
- Ceres seq_id 1568685
- Location of start within SEQ ID NO 1013: at 300 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ank repeat
- Location within SEQ ID NO 1015: from 157 to 190 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1016
- Ceres seq_id 1568686
- Location of start within SEQ ID NO 1013: at 435 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ank repeat
- Location within SEQ ID NO 1016: from 112 to 145 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
142378
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1017
- Ceres seq_id 1568715

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1018
- Ceres seq_id 1568716
- Location of start within SEQ ID NO 1017: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fibronectin type III domain
- Location within SEQ ID NO 1018: from 70 to 156 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1019
- Ceres seq_id 1568717
- Location of start within SEQ ID NO 1017: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fibronectin type III domain
- Location within SEQ ID NO 1019: from 51 to 137 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1020
- Ceres seq_id 1568718
- Location of start within SEQ ID NO 1017: at 542 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
15372
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1021
- Ceres seq_id 1568741
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1022
- Ceres seq_id 1568742
- Location of start within SEQ ID NO 1021: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1023
- Ceres seq_id 1568743
- Location of start within SEQ ID NO 1021: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 493
- gi No. 4926822
- Description: (AC004135) T17H7.7 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 1023: from 1 to 109

Maximum Length Sequence:
related to:
Clone IDs:
142926
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1024
- Ceres seq_id 1568751
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1025
- Ceres seq_id 1568752
- Location of start within SEQ ID NO 1024: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1026
- Ceres seq_id 1568753
- Location of start within SEQ ID NO 1024: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 494
- gi No. 1170089
- Description: GLUTATHIONE S-TRANSFERASE ERD13 (CLASS PHI) >gi|481822|pir||S39542 probable glutathione transferase (EC 2.5.1.18) (clone ERD13) - Arabidopsis thaliana >gi|497789|dbj|BAA04554| (D17673) glutathione S-transferase
- % Identity: 100
- Alignment Length: 48
- Location of Alignment in SEQ ID NO 1026: from 1 to 48

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1027
- Ceres seq_id 1568754
- Location of start within SEQ ID NO 1024: at 240 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
15392
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1028
- Ceres seq_id 1568758
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1029
- Ceres seq_id 1568759
- Location of start within SEQ ID NO 1028: at 215 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 1029: from 187 to 292 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 495
- gi No. 2769566
- Description: (Y10477) chloroplast thylakoidal processing peptidase [Arabidopsis thaliana]
- % Identity: 72.1
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 1029: from 67 to 367

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1030
- Ceres seq_id 1568760
- Location of start within SEQ ID NO 1028: at 437 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 1030: from 113 to 218 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 496
- gi No. 2769566

- Description: (Y10477) chloroplast thylakoidal processing peptidase [Arabidopsis thaliana]
- % Identity: 72.1
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 1030: from 1 to 293

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1031
- Ceres seq_id 1568761
- Location of start within SEQ ID NO 1028: at 500 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 1031: from 92 to 197 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 497
- gi No. 2769566
- Description: (Y10477) chloroplast thylakoidal processing peptidase [Arabidopsis thaliana]
- % Identity: 72.1
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 1031: from 1 to 272

Maximum Length Sequence:
related to:
Clone IDs:
143538
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1032
- Ceres seq_id 1568777

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1033
- Ceres seq_id 1568778
- Location of start within SEQ ID NO 1032: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Catalase
- Location within SEQ ID NO 1033: from 35 to 389 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 498
- gi No. 1705620
- Description: CATALASE 2 >gi|1246399|emb|CAA64220| (X94447) catalase [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 1033: from 18 to 389

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1034
- Ceres seq_id 1568779
- Location of start within SEQ ID NO 1032: at 52 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Catalase
- Location within SEQ ID NO 1034: from 18 to 372 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 499
- gi No. 1705620
- Description: CATALASE 2 >gi|1246399|emb|CAA64220| (X94447) catalase [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 1034: from 1 to 372

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1035
- Ceres seq_id 1568780
- Location of start within SEQ ID NO 1032: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Catalase
- Location within SEQ ID NO 1035: from 1 to 341 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 500
- gi No. 1705620
- Description: CATALASE 2 >gi|1246399|emb|CAA64220| (X94447) catalase [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 372
- Location of Alignment in SEQ ID NO 1035: from 1 to 341

Maximum Length Sequence:
related to:
Clone IDs:
144162
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1036
- Ceres seq_id 1568781

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1037
- Ceres seq_id 1568782
- Location of start within SEQ ID NO 1036: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1038
- Ceres seq_id 1568783
- Location of start within SEQ ID NO 1036: at 14 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1039
- Ceres seq_id 1568784
- Location of start within SEQ ID NO 1036: at 665 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat

- Location within SEQ ID NO 1039: from 82 to 114 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
15640
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1040
- Ceres seq_id 1568803
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1041
- Ceres seq_id 1568804
- Location of start within SEQ ID NO 1040: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AhpC/TSA family
- Location within SEQ ID NO 1041: from 82 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 501
- gi No. 3121825
- Description: 2-CYS PEROXIREDOXIN BAS1 PRECURSOR (THIOL-SPECIFIC ANTIOXIDANT PROTEIN) >gi|1498247|emb|CAA63910| (X94219) bas1 protein [Spinacia oleracea]
- % Identity: 86.7
- Alignment Length: 273
- Location of Alignment in SEQ ID NO 1041: from 3 to 273

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1042
- Ceres seq_id 1568805
- Location of start within SEQ ID NO 1040: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AhpC/TSA family
- Location within SEQ ID NO 1042: from 80 to 231 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 502
- gi No. 3121825
- Description: 2-CYS PEROXIREDOXIN BAS1 PRECURSOR (THIOL-SPECIFIC ANTIOXIDANT PROTEIN) >gi|1498247|emb|CAA63910| (X94219) bas1 protein [Spinacia oleracea]
- % Identity: 86.7
- Alignment Length: 273
- Location of Alignment in SEQ ID NO 1042: from 1 to 271

Maximum Length Sequence:
related to:
Clone IDs:
2944
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1043
- Ceres seq_id 1568826
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1044
- Ceres seq_id 1568827

- Location of start within SEQ ID NO 1043: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 1044: from 196 to 219 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
8291
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1045
- Ceres seq_id 1568859
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1046
- Ceres seq_id 1568860
- Location of start within SEQ ID NO 1045: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type carbonic anhydrases
- Location within SEQ ID NO 1046: from 113 to 288 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1047
- Ceres seq_id 1568861
- Location of start within SEQ ID NO 1045: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type carbonic anhydrases
- Location within SEQ ID NO 1047: from 86 to 261 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1048
- Ceres seq_id 1568862
- Location of start within SEQ ID NO 1045: at 235 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type carbonic anhydrases
- Location within SEQ ID NO 1048: from 35 to 210 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
2102
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1049
- Ceres seq_id 1568863
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1050

- Ceres seq_id 1568864
- Location of start within SEQ ID NO 1049: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 1050: from 1 to 285 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1051
- Ceres seq_id 1568865
- Location of start within SEQ ID NO 1049: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 1051: from 1 to 280 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1052
- Ceres seq_id 1568866
- Location of start within SEQ ID NO 1049: at 327 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridine nucleotide-disulphide oxidoreductase class-I
- Location within SEQ ID NO 1052: from 1 to 177 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
2881
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1053
- Ceres seq_id 1568867

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1054
- Ceres seq_id 1568868
- Location of start within SEQ ID NO 1053: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen allergen
- Location within SEQ ID NO 1054: from 190 to 267 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 503
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.2
- Alignment Length: 232
- Location of Alignment in SEQ ID NO 1054: from 52 to 282

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1055

- Ceres seq_id 1568869
- Location of start within SEQ ID NO 1053: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen allergen
- Location within SEQ ID NO 1055: from 161 to 238 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 504
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.2
- Alignment Length: 232
- Location of Alignment in SEQ ID NO 1055: from 23 to 253

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1056
- Ceres seq_id 1568870
- Location of start within SEQ ID NO 1053: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen allergen
- Location within SEQ ID NO 1056: from 159 to 236 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 505
- gi No. 1041702
- Description: (U30476) expansin At-EXP1 [Arabidopsis thaliana]
- % Identity: 86.2
- Alignment Length: 232
- Location of Alignment in SEQ ID NO 1056: from 21 to 251

Maximum Length Sequence:
related to:
Clone IDs:
15782
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1057
- Ceres seq_id 1568871

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1058
- Ceres seq_id 1568872
- Location of start within SEQ ID NO 1057: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AMP-binding enzyme
- Location within SEQ ID NO 1058: from 1 to 252 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1059
- Ceres seq_id 1568873
- Location of start within SEQ ID NO 1057: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AMP-binding enzyme

- Location within SEQ ID NO 1059: from 1 to 237 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1060
- Ceres seq_id 1568874
- Location of start within SEQ ID NO 1057: at 495 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AMP-binding enzyme
- Location within SEQ ID NO 1060: from 1 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3039
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1061
- Ceres seq_id 1568883
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1062
- Ceres seq_id 1568884
- Location of start within SEQ ID NO 1061: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 1062: from 222 to 306 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 506
- gi No. 3776561
- Description: (AC005388) Identical to DNA for acyl carrier protein (ACP) gene A2 gb|X57699 from A. thaliana. ESTs gb|W43252, gb|T42821, gb|N65229, gb|N97267, gb|F15491 and gb|AA040955 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1062: from 24 to 133

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1063
- Ceres seq_id 1568885
- Location of start within SEQ ID NO 1061: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 1063: from 199 to 283 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 507
- gi No. 3776561
- Description: (AC005388) Identical to DNA for acyl carrier protein (ACP) gene A2 gb|X57699 from A. thaliana. ESTs gb|W43252, gb|T42821, gb|N65229, gb|N97267, gb|F15491 and gb|AA040955 come from this gene. [Arabidopsis thaliana]

- % Identity: 100
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1063: from 1 to 110

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1064
- Ceres seq_id 1568886
- Location of start within SEQ ID NO 1061: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 1064: from 101 to 185 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 508
- gi No. 3776561
- Description: (AC005388) Identical to DNA for acyl carrier protein (ACP) gene A2 gb|X57699 from A. thaliana. ESTs gb|W43252, gb|T42821, gb|N65229, gb|N97267, gb|F15491 and gb|AA040955 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1064: from 1 to 12

Maximum Length Sequence:
related to:
Clone IDs:
5509
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1065
- Ceres seq_id 1568887

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1066
- Ceres seq_id 1568888
- Location of start within SEQ ID NO 1065: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 509
- gi No. 4982522
- Description: (AC000107) F17F8.28 [Arabidopsis thaliana]
- % Identity: 78.3
- Alignment Length: 139
- Location of Alignment in SEQ ID NO 1066: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1067
- Ceres seq_id 1568889
- Location of start within SEQ ID NO 1065: at 425 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 510
- gi No. 4982522
- Description: (AC000107) F17F8.28 [Arabidopsis thaliana]
- % Identity: 78.3
- Alignment Length: 139
- Location of Alignment in SEQ ID NO 1067: from 1 to 20

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1068
- Ceres seq_id 1568890
- Location of start within SEQ ID NO 1065: at 675 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 511
- gi No. 4982522
- Description: (AC000107) F17F8.28 [Arabidopsis thaliana]
- % Identity: 91.8
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 1068: from 23 to 180

Maximum Length Sequence:
related to:
Clone IDs:
799

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1069
- Ceres seq_id 1568891

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1070
- Ceres seq_id 1568892
- Location of start within SEQ ID NO 1069: at 283 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- pfkB family carbohydrate kinase
- Location within SEQ ID NO 1070: from 91 to 324 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 512
- gi No. 4469011
- Description: (AL035602) carbohydrate kinase-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 1070: from 1 to 385

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1071
- Ceres seq_id 1568893
- Location of start within SEQ ID NO 1069: at 412 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- pfkB family carbohydrate kinase
- Location within SEQ ID NO 1071: from 48 to 281 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 513
- gi No. 4469011
- Description: (AL035602) carbohydrate kinase-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 1071: from 1 to 342

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1072
- Ceres seq_id 1568894
- Location of start within SEQ ID NO 1069: at 523 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- pfkB family carbohydrate kinase
- Location within SEQ ID NO 1072: from 11 to 244 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 514
- gi No. 4469011
- Description: (AL035602) carbohydrate kinase-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 1072: from 1 to 305

Maximum Length Sequence:
related to:
Clone IDs:
639
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1073
- Ceres seq_id 1568895
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1074
- Ceres seq_id 1568896
- Location of start within SEQ ID NO 1073: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enoyl-CoA hydratase/isomerase family
- Location within SEQ ID NO 1074: from 20 to 191 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1075
- Ceres seq_id 1568897
- Location of start within SEQ ID NO 1073: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enoyl-CoA hydratase/isomerase family
- Location within SEQ ID NO 1075: from 16 to 187 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1076
- Ceres seq_id 1568898
- Location of start within SEQ ID NO 1073: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enoyl-CoA hydratase/isomerase family
- Location within SEQ ID NO 1076: from 1 to 155 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6723
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1077
- Ceres seq_id 1568902
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1078
- Ceres seq_id 1568903
- Location of start within SEQ ID NO 1077: at 23 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1078: from 108 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 515
- gi No. 2244985
- Description: (Z97340) peroxidase like protein [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 121
- Location of Alignment in SEQ ID NO 1078: from 108 to 228

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1079
- Ceres seq_id 1568904
- Location of start within SEQ ID NO 1077: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1079: from 94 to 214 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 516
- gi No. 2244985
- Description: (Z97340) peroxidase like protein [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 121
- Location of Alignment in SEQ ID NO 1079: from 94 to 214

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1080
- Ceres seq_id 1568905
- Location of start within SEQ ID NO 1077: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1080: from 85 to 205 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 517
- gi No. 2244985
- Description: (Z97340) peroxidase like protein [Arabidopsis thaliana]

- % Identity: 98.3
- Alignment Length: 121
- Location of Alignment in SEQ ID NO 1080: from 85 to 205

Maximum Length Sequence:
related to:
Clone IDs:
2641
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1081
- Ceres seq_id 1568906
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1082
- Ceres seq_id 1568907
- Location of start within SEQ ID NO 1081: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1082: from 10 to 182 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1083
- Ceres seq_id 1568908
- Location of start within SEQ ID NO 1081: at 469 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6568
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1084
- Ceres seq_id 1568928
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1085
- Ceres seq_id 1568929
- Location of start within SEQ ID NO 1084: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 1085: from 218 to 282 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1086
- Ceres seq_id 1568930
- Location of start within SEQ ID NO 1084: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 1086: from 213 to 277 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1087
- Ceres seq_id 1568931
- Location of start within SEQ ID NO 1084: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- bZIP transcription factor
- Location within SEQ ID NO 1087: from 168 to 232 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
7559
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1088
- Ceres seq_id 1568944
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1089
- Ceres seq_id 1568945
- Location of start within SEQ ID NO 1088: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1089: from 9 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 518
- gi No. 4775270
- Description: (AJ131214) SF2/ASF-like splicing modulator Srp30 [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 270
- Location of Alignment in SEQ ID NO 1089: from 1 to 270

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1090
- Ceres seq_id 1568946
- Location of start within SEQ ID NO 1088: at 518 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 519
- gi No. 4775270
- Description: (AJ131214) SF2/ASF-like splicing modulator Srp30 [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 270
- Location of Alignment in SEQ ID NO 1090: from 1 to 142

Maximum Length Sequence:
related to:
Clone IDs:
20858
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 1091
- Ceres seq_id 1568962

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1092
- Ceres seq_id 1568963
- Location of start within SEQ ID NO 1091: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 520
- gi No. 2213592
- Description: (AC000348) T7N9.12 [Arabidopsis thaliana]
- % Identity: 96.9
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 1092: from 28 to 251

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1093
- Ceres seq_id 1568964
- Location of start within SEQ ID NO 1091: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 521
- gi No. 2213592
- Description: (AC000348) T7N9.12 [Arabidopsis thaliana]
- % Identity: 96.9
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 1093: from 1 to 206

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1094
- Ceres seq_id 1568965
- Location of start within SEQ ID NO 1091: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 522
- gi No. 2213592
- Description: (AC000348) T7N9.12 [Arabidopsis thaliana]
- % Identity: 96.9
- Alignment Length: 224
- Location of Alignment in SEQ ID NO 1094: from 1 to 195

Maximum Length Sequence:
related to:
Clone IDs:
25530

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1095
- Ceres seq_id 1568966

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1096
- Ceres seq_id 1568967
- Location of start within SEQ ID NO 1095: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- ABC transporter
- Location within SEQ ID NO 1096: from 55 to 233 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1097
- Ceres seq_id 1568968
- Location of start within SEQ ID NO 1095: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 1097: from 1 to 178 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1098
- Ceres seq_id 1568969
- Location of start within SEQ ID NO 1095: at 454 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 1098: from 1 to 127 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
19080
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1099
- Ceres seq_id 1568974

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1100
- Ceres seq_id 1568975
- Location of start within SEQ ID NO 1099: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 523
- gi No. 5059396
- Description: (AF155817) zinc finger protein Dof3 [Arabidopsis thaliana]
- % Identity: 91.9
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 1100: from 1 to 284

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1101
- Ceres seq_id 1568976
- Location of start within SEQ ID NO 1099: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 524

- gi No. 5059396
- Description: (AF155817) zinc finger protein Dof3 [Arabidopsis thaliana]
- % Identity: 91.9
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 1101: from 1 to 276

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1102
- Ceres seq_id 1568977
- Location of start within SEQ ID NO 1099: at 223 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 525
- gi No. 5059396
- Description: (AF155817) zinc finger protein Dof3 [Arabidopsis thaliana]
- % Identity: 91.9
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 1102: from 1 to 248

Maximum Length Sequence:
related to:
Clone IDs:
18922

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1103
- Ceres seq_id 1568978

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1104
- Ceres seq_id 1568979
- Location of start within SEQ ID NO 1103: at 278 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Shikimate kinase
- Location within SEQ ID NO 1104: from 103 to 157 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1105
- Ceres seq_id 1568980
- Location of start within SEQ ID NO 1103: at 696 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Shikimate kinase
- Location within SEQ ID NO 1105: from 1 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 526
- gi No. 5042170
- Description: (AL078620) shikimate kinase-like protein [Arabidopsis thaliana]
- % Identity: 78.1
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 1105: from 1 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1106
- Ceres seq_id 1568981
- Location of start within SEQ ID NO 1103: at 852 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Shikimate kinase
- Location within SEQ ID NO 1106: from 1 to 94 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 527
- gi No. 5042170
- Description: (AL078620) shikimate kinase-like protein [Arabidopsis thaliana]
- % Identity: 78.1
- Alignment Length: 169
- Location of Alignment in SEQ ID NO 1106: from 1 to 109

Maximum Length Sequence:
related to:
Clone IDs:
22343
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1107
- Ceres seq_id 1569006
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1108
- Ceres seq_id 1569007
- Location of start within SEQ ID NO 1107: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 1108: from 75 to 123 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1109
- Ceres seq_id 1569008
- Location of start within SEQ ID NO 1107: at 484 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1110
- Ceres seq_id 1569009
- Location of start within SEQ ID NO 1107: at 562 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20888
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 1111
- Ceres seq_id 1569025

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1112
- Ceres seq_id 1569026
- Location of start within SEQ ID NO 1111: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Serine carboxypeptidase
- Location within SEQ ID NO 1112: from 49 to 465 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 528
- gi No. 2980785
- Description: (AL022198) SERINE CARBOXYPEPTIDASE II-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 190
- Location of Alignment in SEQ ID NO 1112: from 8 to 197

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1113
- Ceres seq_id 1569027
- Location of start within SEQ ID NO 1111: at 24 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Serine carboxypeptidase
- Location within SEQ ID NO 1113: from 42 to 458 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 529
- gi No. 2980785
- Description: (AL022198) SERINE CARBOXYPEPTIDASE II-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 190
- Location of Alignment in SEQ ID NO 1113: from 1 to 190

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1114
- Ceres seq_id 1569028
- Location of start within SEQ ID NO 1111: at 672 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Serine carboxypeptidase
- Location within SEQ ID NO 1114: from 1 to 242 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 530
- gi No. 2980785
- Description: (AL022198) SERINE CARBOXYPEPTIDASE II-like protein [Arabidopsis thaliana]
- % Identity: 92.9
- Alignment Length: 168
- Location of Alignment in SEQ ID NO 1114: from 96 to 263

Maximum Length Sequence:
related to:

Clone IDs:
20915
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1115
- Ceres seq_id 1569029
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1116
- Ceres seq_id 1569030
- Location of start within SEQ ID NO 1115: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 531
- gi No. 4678941
- Description: (AL049711) gamma response I protein [Arabidopsis thaliana]
- % Identity: 93.2
- Alignment Length: 497
- Location of Alignment in SEQ ID NO 1116: from 9 to 486

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1117
- Ceres seq_id 1569031
- Location of start within SEQ ID NO 1115: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 532
- gi No. 4678941
- Description: (AL049711) gamma response I protein [Arabidopsis thaliana]
- % Identity: 93.2
- Alignment Length: 497
- Location of Alignment in SEQ ID NO 1117: from 1 to 443

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1118
- Ceres seq_id 1569032
- Location of start within SEQ ID NO 1115: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 533
- gi No. 4678941
- Description: (AL049711) gamma response I protein [Arabidopsis thaliana]
- % Identity: 93.2
- Alignment Length: 497
- Location of Alignment in SEQ ID NO 1118: from 1 to 427

Maximum Length Sequence:
related to:
Clone IDs:
21241
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1119
- Ceres seq_id 1569041
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1120
- Ceres seq_id 1569042
- Location of start within SEQ ID NO 1119: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1120: from 16 to 249 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1121
- Ceres seq_id 1569043
- Location of start within SEQ ID NO 1119: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1121: from 7 to 240 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1122
- Ceres seq_id 1569044
- Location of start within SEQ ID NO 1119: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1122: from 1 to 167 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20877
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1123
- Ceres seq_id 1569049

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1124
- Ceres seq_id 1569050
- Location of start within SEQ ID NO 1123: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1124: from 95 to 234 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1125
- Ceres seq_id 1569051
- Location of start within SEQ ID NO 1123: at 250 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1125: from 44 to 183 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1126
- Ceres seq_id 1569052
- Location of start within SEQ ID NO 1123: at 445 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1126: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
7575
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1127
- Ceres seq_id 1569061

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1128
- Ceres seq_id 1569062
- Location of start within SEQ ID NO 1127: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1128: from 70 to 329 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1129
- Ceres seq_id 1569063
- Location of start within SEQ ID NO 1127: at 314 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1129: from 1 to 225 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1130
- Ceres seq_id 1569064
- Location of start within SEQ ID NO 1127: at 341 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1130: from 1 to 216 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

related to:
Clone IDs:
21250
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1131
- Ceres seq_id 1569065
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1132
- Ceres seq_id 1569066
- Location of start within SEQ ID NO 1131: at 668 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1132: from 1 to 195 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1133
- Ceres seq_id 1569067
- Location of start within SEQ ID NO 1131: at 743 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1133: from 1 to 170 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1134
- Ceres seq_id 1569068
- Location of start within SEQ ID NO 1131: at 893 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1134: from 1 to 120 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
1360
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1135
- Ceres seq_id 1569097
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1136
- Ceres seq_id 1569098
- Location of start within SEQ ID NO 1135: at 306 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 534
- gi No. 3319350
- Description: (AF077407) No definition line found [Arabidopsis thaliana]

- % Identity: 85.8
- Alignment Length: 232
- Location of Alignment in SEQ ID NO 1136: from 21 to 252

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1137
- Ceres seq_id 1569099
- Location of start within SEQ ID NO 1135: at 330 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 535
- gi No. 3319350
- Description: (AF077407) No definition line found [Arabidopsis thaliana]
- % Identity: 85.8
- Alignment Length: 232
- Location of Alignment in SEQ ID NO 1137: from 13 to 244

Maximum Length Sequence:
related to:
Clone IDs:
28640
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1138
- Ceres seq_id 1569112

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1139
- Ceres seq_id 1569113
- Location of start within SEQ ID NO 1138: at 529 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1140
- Ceres seq_id 1569114
- Location of start within SEQ ID NO 1138: at 1017 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 536
- gi No. 2191135
- Description: (AF007269) A_IG002N01.14 gene product [Arabidopsis thaliana]
- % Identity: 95.8
- Alignment Length: 383
- Location of Alignment in SEQ ID NO 1140: from 1 to 109

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1141
- Ceres seq_id 1569115
- Location of start within SEQ ID NO 1138: at 1044 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 537

- gi No. 2191135
- Description: (AF007269) A_IG002N01.14 gene product [Arabidopsis thaliana]
- % Identity: 95.8
- Alignment Length: 383
- Location of Alignment in SEQ ID NO 1141: from 1 to 100

Maximum Length Sequence:
related to:
Clone IDs:
37277
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1142
- Ceres seq_id 1569116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1143
- Ceres seq_id 1569117
- Location of start within SEQ ID NO 1142: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sigma-70 factor
- Location within SEQ ID NO 1143: from 287 to 518 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 538
- gi No. 2353171
- Description: (AF015542) sigma factor 1 [Arabidopsis thaliana] >gi|2443408|dbj|BAA22421| (D89993) SigA [Arabidopsis thaliana] >gi|2558514|emb|CAA74640| (Y14252) plastid RNA polymerase sigma factor [Arabidopsis thaliana] >gi|5042421|gb|AAD38260.1|AC006193_16
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1143: from 31 to 532

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1144
- Ceres seq_id 1569118
- Location of start within SEQ ID NO 1142: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sigma-70 factor
- Location within SEQ ID NO 1144: from 257 to 488 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 539
- gi No. 2353171
- Description: (AF015542) sigma factor 1 [Arabidopsis thaliana] >gi|2443408|dbj|BAA22421| (D89993) SigA [Arabidopsis thaliana] >gi|2558514|emb|CAA74640| (Y14252) plastid RNA polymerase sigma factor [Arabidopsis thaliana] >gi|5042421|gb|AAD38260.1|AC006193_16
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1144: from 1 to 502

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1145
- Ceres seq_id 1569119
- Location of start within SEQ ID NO 1142: at 452 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sigma-70 factor
- Location within SEQ ID NO 1145: from 137 to 368 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 540
- gi No. 2353171
- Description: (AF015542) sigma factor 1 [Arabidopsis thaliana] >gi|2443408|dbj|BAA22421| (D89993) SigA [Arabidopsis thaliana] >gi|2558514|emb|CAA74640| (Y14252) plastid RNA polymerase sigma factor [Arabidopsis thaliana] >gi|5042421|gb|AAD38260.1|AC006193_16
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1145: from 1 to 382

Maximum Length Sequence:
related to:
Clone IDs:
34432
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1146
- Ceres seq_id 1569144
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1147
- Ceres seq_id 1569145
- Location of start within SEQ ID NO 1146: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 1147: from 48 to 93 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1148
- Ceres seq_id 1569146
- Location of start within SEQ ID NO 1146: at 448 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1149
- Ceres seq_id 1569147
- Location of start within SEQ ID NO 1146: at 466 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34700
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1150
- Ceres seq_id 1569148
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1151
- Ceres seq_id 1569149
- Location of start within SEQ ID NO 1150: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial transferase hexapeptide (four repeats)
- Location within SEQ ID NO 1151: from 59 to 88 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1152
- Ceres seq_id 1569150
- Location of start within SEQ ID NO 1150: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial transferase hexapeptide (four repeats)
- Location within SEQ ID NO 1152: from 56 to 85 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1153
- Ceres seq_id 1569151
- Location of start within SEQ ID NO 1150: at 271 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial transferase hexapeptide (four repeats)
- Location within SEQ ID NO 1153: from 14 to 43 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34867
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1154
- Ceres seq_id 1569156

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1155
- Ceres seq_id 1569157
- Location of start within SEQ ID NO 1154: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1155: from 53 to 299 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 541
- gi No. 2832707
- Description: (AL021713) translation initiation factor eIF-2 gamma chain-like protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1155: from 17 to 291

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1156
- Ceres seq_id 1569158
- Location of start within SEQ ID NO 1154: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1156: from 37 to 283 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 542
- gi No. 2832707
- Description: (AL021713) translation initiation factor eIF-2 gamma chain-like protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1156: from 1 to 275

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1157
- Ceres seq_id 1569159
- Location of start within SEQ ID NO 1154: at 341 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1157: from 1 to 186 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 543
- gi No. 2832707
- Description: (AL021713) translation initiation factor eIF-2 gamma chain-like protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1157: from 1 to 178

Maximum Length Sequence:
related to:
Clone IDs:
36467
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1158
- Ceres seq_id 1569164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1159
- Ceres seq_id 1569165
- Location of start within SEQ ID NO 1158: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 1159: from 60 to 290 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1160
- Ceres seq_id 1569166
- Location of start within SEQ ID NO 1158: at 346 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 1160: from 1 to 230 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1161
- Ceres seq_id 1569167
- Location of start within SEQ ID NO 1158: at 547 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 1161: from 1 to 163 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39051
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1162
- Ceres seq_id 1569168
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1163
- Ceres seq_id 1569169
- Location of start within SEQ ID NO 1162: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1163: from 5 to 405 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 544
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 431
- Location of Alignment in SEQ ID NO 1163: from 1 to 430

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1164
- Ceres seq_id 1569170
- Location of start within SEQ ID NO 1162: at 238 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1164: from 1 to 357 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 545
- gi No. 119143

- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 431
- Location of Alignment in SEQ ID NO 1164: from 1 to 382

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1165
- Ceres seq_id 1569171
- Location of start within SEQ ID NO 1162: at 397 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1165: from 1 to 304 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 546
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 431
- Location of Alignment in SEQ ID NO 1165: from 1 to 329

Maximum Length Sequence:
related to:
Clone IDs:
21819
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1166
- Ceres seq_id 1569179

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1167
- Ceres seq_id 1569180
- Location of start within SEQ ID NO 1166: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1168
- Ceres seq_id 1569181
- Location of start within SEQ ID NO 1166: at 444 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rieske iron-sulfur protein
- Location within SEQ ID NO 1168: from 41 to 164 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 547
- gi No. 1076675
- Description: ubiquinol--cytochrome-c reductase (EC 1.10.2.2) iron-sulfur protein - potato
- % Identity: 88.7

- Alignment Length: 151
- Location of Alignment in SEQ ID NO 1168: from 30 to 180

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1169
- Ceres seq_id 1569182
- Location of start within SEQ ID NO 1166: at 459 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rieske iron-sulfur protein
- Location within SEQ ID NO 1169: from 36 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 548
- gi No. 1076675
- Description: ubiquinol--cytochrome-c reductase (EC 1.10.2.2) iron-sulfur protein - potato
- % Identity: 88.7
- Alignment Length: 151
- Location of Alignment in SEQ ID NO 1169: from 25 to 175

Maximum Length Sequence:
related to:
Clone IDs:
13864
20829
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1170
- Ceres seq_id 1569183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1171
- Ceres seq_id 1569184
- Location of start within SEQ ID NO 1170: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1171: from 62 to 238 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1172
- Ceres seq_id 1569185
- Location of start within SEQ ID NO 1170: at 561 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1172: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1173
- Ceres seq_id 1569186
- Location of start within SEQ ID NO 1170: at 702 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
25128
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1174
- Ceres seq_id 1569187
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1175
- Ceres seq_id 1569188
- Location of start within SEQ ID NO 1174: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2-oxo acid dehydrogenases acyltransferase (catalytic domain)
- Location within SEQ ID NO 1175: from 287 to 507 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 549
- gi No. 3138972
- Description: (AF038505) dihydrolipoylacyltransferase subunit of the branched-chain alpha-keto acid dehydrogenase complex [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 483
- Location of Alignment in SEQ ID NO 1175: from 27 to 509

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1176
- Ceres seq_id 1569189
- Location of start within SEQ ID NO 1174: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2-oxo acid dehydrogenases acyltransferase (catalytic domain)
- Location within SEQ ID NO 1176: from 261 to 481 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 550
- gi No. 3138972
- Description: (AF038505) dihydrolipoylacyltransferase subunit of the branched-chain alpha-keto acid dehydrogenase complex [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 483
- Location of Alignment in SEQ ID NO 1176: from 1 to 483

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1177
- Ceres seq_id 1569190
- Location of start within SEQ ID NO 1174: at 232 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2-oxo acid dehydrogenases acyltransferase (catalytic domain)
- Location within SEQ ID NO 1177: from 210 to 430 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 551

- gi No. 3138972
- Description: (AF038505) dihydrolipoylacyltransferase subunit of the branched-chain alpha-keto acid dehydrogenase complex [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 483
- Location of Alignment in SEQ ID NO 1177: from 1 to 432

Maximum Length Sequence:
related to:
Clone IDs:
16423
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1178
- Ceres seq_id 1569230
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1179
- Ceres seq_id 1569231
- Location of start within SEQ ID NO 1178: at 225 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 552
- gi No. 1360088
- Description: (X97947) Zn finger protein [Nicotiana tabacum]
- % Identity: 79.4
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 1179: from 21 to 83

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1180
- Ceres seq_id 1569232
- Location of start within SEQ ID NO 1178: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 553
- gi No. 1360088
- Description: (X97947) Zn finger protein [Nicotiana tabacum]
- % Identity: 79.4
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 1180: from 17 to 79

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1181
- Ceres seq_id 1569233
- Location of start within SEQ ID NO 1178: at 531 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
17434
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1182
- Ceres seq_id 1569234

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1183
- Ceres seq_id 1569235
- Location of start within SEQ ID NO 1182: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 554
- gi No. 3341723
- Description: (AF052690) CONSTANS-like 1 protein [Raphanus sativus]
- % Identity: 73.2
- Alignment Length: 313
- Location of Alignment in SEQ ID NO 1183: from 4 to 294

Maximum Length Sequence:
related to:
Clone IDs:
37175
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1184
- Ceres seq_id 1569255
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1185
- Ceres seq_id 1569256
- Location of start within SEQ ID NO 1184: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 555
- gi No. 2252858
- Description: (AF013294) A_TM018A10.11 gene product [Arabidopsis thaliana]
- % Identity: 88.5
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 1185: from 240 to 346

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1186
- Ceres seq_id 1569257
- Location of start within SEQ ID NO 1184: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 556
- gi No. 2252858
- Description: (AF013294) A_TM018A10.11 gene product [Arabidopsis thaliana]
- % Identity: 88.5
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 1186: from 234 to 340

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1187
- Ceres seq_id 1569258
- Location of start within SEQ ID NO 1184: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 557
- gi No. 2252858
- Description: (AF013294) A_TM018A10.11 gene product [Arabidopsis thaliana]
- % Identity: 88.5
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 1187: from 203 to 309

Maximum Length Sequence:
related to:
Clone IDs:
37111
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1188
- Ceres seq_id 1569277
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1189
- Ceres seq_id 1569278
- Location of start within SEQ ID NO 1188: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 1189: from 66 to 325 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34582
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1190
- Ceres seq_id 1569286
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1191
- Ceres seq_id 1569287
- Location of start within SEQ ID NO 1190: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Armadillo/beta-catenin-like repeats
- Location within SEQ ID NO 1191: from 177 to 215 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1192
- Ceres seq_id 1569288
- Location of start within SEQ ID NO 1190: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Armadillo/beta-catenin-like repeats
- Location within SEQ ID NO 1192: from 175 to 213 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1193
- Ceres seq_id 1569289
- Location of start within SEQ ID NO 1190: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Armadillo/beta-catenin-like repeats
- Location within SEQ ID NO 1193: from 164 to 202 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37733
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1194
- Ceres seq_id 1569294

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1195
- Ceres seq_id 1569295
- Location of start within SEQ ID NO 1194: at 252 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Squalene and phytoene synthases
- Location within SEQ ID NO 1195: from 1 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 558
- gi No. 2228795
- Description: (AF004396) squalene synthase 2 [Arabidopsis thaliana]
- % Identity: 90.5
- Alignment Length: 233
- Location of Alignment in SEQ ID NO 1195: from 1 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1196
- Ceres seq_id 1569296
- Location of start within SEQ ID NO 1194: at 336 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Squalene and phytoene synthases
- Location within SEQ ID NO 1196: from 1 to 131 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 559
- gi No. 2228795
- Description: (AF004396) squalene synthase 2 [Arabidopsis thaliana]
- % Identity: 90.5
- Alignment Length: 233
- Location of Alignment in SEQ ID NO 1196: from 1 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1197
- Ceres seq_id 1569297
- Location of start within SEQ ID NO 1194: at 348 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Squalene and phytoene synthases
- Location within SEQ ID NO 1197: from 1 to 127 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 560
- gi No. 2228795
- Description: (AF004396) squalene synthase 2 [Arabidopsis thaliana]
- % Identity: 90.5
- Alignment Length: 233
- Location of Alignment in SEQ ID NO 1197: from 1 to 127

Maximum Length Sequence:
related to:
Clone IDs:
32761
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1198
- Ceres seq_id 1569298
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1199
- Ceres seq_id 1569299
- Location of start within SEQ ID NO 1198: at 213 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1199: from 27 to 246 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1200
- Ceres seq_id 1569300
- Location of start within SEQ ID NO 1198: at 246 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1200: from 16 to 235 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1201
- Ceres seq_id 1569301
- Location of start within SEQ ID NO 1198: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1201: from 9 to 228 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

33972
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1202
- Ceres seq_id 1569306
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1203
- Ceres seq_id 1569307
- Location of start within SEQ ID NO 1202: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1203: from 22 to 278 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 561
- gi No. 2146746
- Description: protein kinase (EC 2.7.1.-) - Arabidopsis thaliana >gi|166819 (L05562) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 361
- Location of Alignment in SEQ ID NO 1203: from 1 to 361

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1204
- Ceres seq_id 1569308
- Location of start within SEQ ID NO 1202: at 151 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1204: from 10 to 266 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 562
- gi No. 2146746
- Description: protein kinase (EC 2.7.1.-) - Arabidopsis thaliana >gi|166819 (L05562) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 361
- Location of Alignment in SEQ ID NO 1204: from 1 to 349

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1205
- Ceres seq_id 1569309
- Location of start within SEQ ID NO 1202: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1205: from 7 to 263 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 563
- gi No. 2146746
- Description: protein kinase (EC 2.7.1.-) - Arabidopsis thaliana >gi|166819 (L05562) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 361
- Location of Alignment in SEQ ID NO 1205: from 1 to 346

Maximum Length Sequence:
related to:
Clone IDs:
34360
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1206
- Ceres seq_id 1569310
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1207
- Ceres seq_id 1569311
- Location of start within SEQ ID NO 1206: at 54 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 1207: from 30 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 564
- gi No. 1168256
- Description: ASPARTATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR (TRANSAMINASE A) >gi|693688 (U15026) aspartate aminotransferase [Arabidopsis thaliana] >gi|3201622 (AC004669) aspartate aminotransferase [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 1207: from 1 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1208
- Ceres seq_id 1569312
- Location of start within SEQ ID NO 1206: at 629 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 1208: from 1 to 236 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 565
- gi No. 1168256
- Description: ASPARTATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR (TRANSAMINASE A) >gi|693688 (U15026) aspartate aminotransferase [Arabidopsis thaliana] >gi|3201622 (AC004669) aspartate aminotransferase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 322
- Location of Alignment in SEQ ID NO 1208: from 1 to 238

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1209
- Ceres seq_id 1569313
- Location of start within SEQ ID NO 1206: at 776 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 1209: from 1 to 187 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 566

- gi No. 1168256
- Description: ASPARTATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR (TRANSAMINASE A) >gi|693688 (U15026) aspartate aminotransferase [Arabidopsis thaliana] >gi|3201622 (AC004669) aspartate aminotransferase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 322
- Location of Alignment in SEQ ID NO 1209: from 1 to 189

Maximum Length Sequence:
related to:
Clone IDs:
38858
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1210
- Ceres seq_id 1569326
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1211
- Ceres seq_id 1569327
- Location of start within SEQ ID NO 1210: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 567
- gi No. 2443890
- Description: (AC002294) similar to NAM (gp|X92205|1321924) and CUC2 (gp|AB002560|1944132) proteins [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 184
- Location of Alignment in SEQ ID NO 1211: from 4 to 187

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1212
- Ceres seq_id 1569328
- Location of start within SEQ ID NO 1210: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 568
- gi No. 2443890
- Description: (AC002294) similar to NAM (gp|X92205|1321924) and CUC2 (gp|AB002560|1944132) proteins [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 184
- Location of Alignment in SEQ ID NO 1212: from 1 to 184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1213
- Ceres seq_id 1569329
- Location of start within SEQ ID NO 1210: at 705 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 569
- gi No. 2443890
- Description: (AC002294) similar to NAM (gp|X92205|1321924) and CUC2 (gp|AB002560|1944132) proteins [Arabidopsis thaliana]
- % Identity: 94.9

- Alignment Length: 78
- Location of Alignment in SEQ ID NO 1213: from 1 to 73

Maximum Length Sequence:
related to:
Clone IDs:
147819
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1214
- Ceres seq_id 1569338
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1215
- Ceres seq_id 1569339
- Location of start within SEQ ID NO 1214: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1215: from 64 to 135 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1216
- Ceres seq_id 1569340
- Location of start within SEQ ID NO 1214: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1216: from 38 to 109 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1217
- Ceres seq_id 1569341
- Location of start within SEQ ID NO 1214: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1217: from 35 to 106 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
33557
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1218
- Ceres seq_id 1569354
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1219
- Ceres seq_id 1569355
- Location of start within SEQ ID NO 1218: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins

- Location within SEQ ID NO 1219: from 22 to 129 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 570
- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 129
- Location of Alignment in SEQ ID NO 1219: from 1 to 129

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1220
- Ceres seq_id 1569356
- Location of start within SEQ ID NO 1218: at 461 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 1220: from 2 to 170 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 571
- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 177
- Location of Alignment in SEQ ID NO 1220: from 2 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1221
- Ceres seq_id 1569357
- Location of start within SEQ ID NO 1218: at 503 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 1221: from 1 to 156 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 572
- gi No. 5080790
- Description: (AC007576) Similar to mitochondrial carrier proteins [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 177
- Location of Alignment in SEQ ID NO 1221: from 1 to 164

Maximum Length Sequence:
related to:
Clone IDs:
147869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1222
- Ceres seq_id 1569358

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1223
- Ceres seq_id 1569359
- Location of start within SEQ ID NO 1222: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chorismate synthase
- Location within SEQ ID NO 1223: from 101 to 451 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 573
- gi No. 2492952
- Description: CHORISMATE SYNTHASE 1 PRECURSOR (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE 1) >gi|542026|pir||S40410 chorismate synthase (EC 4.6.1.4) 1 precursor - tomato >gi|410482|emb|CAA79859| (Z21796) chorismate synthase 1 [Lycopersicon esculentum]
- % Identity: 79.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 1223: from 49 to 475

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1224
- Ceres seq_id 1569360
- Location of start within SEQ ID NO 1222: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chorismate synthase
- Location within SEQ ID NO 1224: from 58 to 408 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 574
- gi No. 2492952
- Description: CHORISMATE SYNTHASE 1 PRECURSOR (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE 1) >gi|542026|pir||S40410 chorismate synthase (EC 4.6.1.4) 1 precursor - tomato >gi|410482|emb|CAA79859| (Z21796) chorismate synthase 1 [Lycopersicon esculentum]
- % Identity: 79.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 1224: from 6 to 432

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1225
- Ceres seq_id 1569361
- Location of start within SEQ ID NO 1222: at 497 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chorismate synthase
- Location within SEQ ID NO 1225: from 1 to 286 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 575
- gi No. 2492952
- Description: CHORISMATE SYNTHASE 1 PRECURSOR (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE 1) >gi|542026|pir||S40410 chorismate synthase (EC 4.6.1.4) 1 precursor - tomato >gi|410482|emb|CAA79859| (Z21796) chorismate synthase 1 [Lycopersicon esculentum]
- % Identity: 79.2
- Alignment Length: 427
- Location of Alignment in SEQ ID NO 1225: from 1 to 310

Maximum Length Sequence:
related to:
Clone IDs:
148102
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1226
- Ceres seq_id 1569370
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1227
- Ceres seq_id 1569371
- Location of start within SEQ ID NO 1226: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1228
- Ceres seq_id 1569372
- Location of start within SEQ ID NO 1226: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1229
- Ceres seq_id 1569373
- Location of start within SEQ ID NO 1226: at 423 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1229: from 51 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 576
- gi No. 1946988
- Description: (AF000193) T20B6.3 gene product [Caenorhabditis elegans]
- % Identity: 70.8
- Alignment Length: 93
- Location of Alignment in SEQ ID NO 1229: from 49 to 137

Maximum Length Sequence:
related to:
Clone IDs:
148597
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1230
- Ceres seq_id 1569378
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1231
- Ceres seq_id 1569379
- Location of start within SEQ ID NO 1230: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 577
- gi No. 4973258

- Description: (AF144388) thioredoxin-like 2 [Arabidopsis thaliana]
- % Identity: 73.2
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 1231: from 15 to 101

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1232
- Ceres seq_id 1569380
- Location of start within SEQ ID NO 1230: at 23 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 578
- gi No. 4973258
- Description: (AF144388) thioredoxin-like 2 [Arabidopsis thaliana]
- % Identity: 73.2
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 1232: from 8 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1233
- Ceres seq_id 1569381
- Location of start within SEQ ID NO 1230: at 304 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1233: from 19 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 579
- gi No. 4973258
- Description: (AF144388) thioredoxin-like 2 [Arabidopsis thaliana]
- % Identity: 93.5
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 1233: from 1 to 127

Maximum Length Sequence:
related to:
Clone IDs:
149102

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1234
- Ceres seq_id 1569385

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1235
- Ceres seq_id 1569386
- Location of start within SEQ ID NO 1234: at 569 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 580
- gi No. 4455217
- Description: (AL035440) Avr9 elicitor response like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 332
- Location of Alignment in SEQ ID NO 1235: from 1 to 289

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1236
- Ceres seq_id 1569387
- Location of start within SEQ ID NO 1234: at 635 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 581
- gi No. 4455217
- Description: (AL035440) Avr9 elicitor response like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 332
- Location of Alignment in SEQ ID NO 1236: from 1 to 267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1237
- Ceres seq_id 1569388
- Location of start within SEQ ID NO 1234: at 701 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 582
- gi No. 4455217
- Description: (AL035440) Avr9 elicitor response like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 332
- Location of Alignment in SEQ ID NO 1237: from 1 to 245

Maximum Length Sequence:
related to:
Clone IDs:
149466
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1238
- Ceres seq_id 1569392

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1239
- Ceres seq_id 1569393
- Location of start within SEQ ID NO 1238: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 583
- gi No. 2435518
- Description: (AF024504) contains similarity to C3HC4-type zinc fingers [Arabidopsis thaliana]
- % Identity: 70.9
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 1239: from 43 to 117

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1240
- Ceres seq_id 1569394
- Location of start within SEQ ID NO 1238: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 584
- gi No. 2435518
- Description: (AF024504) contains similarity to C3HC4-type zinc fingers [Arabidopsis thaliana]
- % Identity: 70.9
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 1240: from 19 to 93

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1241
- Ceres seq_id 1569395
- Location of start within SEQ ID NO 1238: at 320 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1241: from 79 to 120 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
18363
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1242
- Ceres seq_id 1569403

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1243
- Ceres seq_id 1569404
- Location of start within SEQ ID NO 1242: at 168 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 10
- Location within SEQ ID NO 1243: from 251 to 336 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1244
- Ceres seq_id 1569405
- Location of start within SEQ ID NO 1242: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 10
- Location within SEQ ID NO 1244: from 246 to 331 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1245
- Ceres seq_id 1569406
- Location of start within SEQ ID NO 1242: at 330 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Glycosyl hydrolase family 10
- Location within SEQ ID NO 1245: from 197 to 282 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
25257
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1246
- Ceres seq_id 1569407
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1247
- Ceres seq_id 1569408
- Location of start within SEQ ID NO 1246: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase family
- Location within SEQ ID NO 1247: from 37 to 284 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 585
- gi No. 4538897
- Description: (AL049482) AX110P-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 362
- Location of Alignment in SEQ ID NO 1247: from 31 to 392

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1248
- Ceres seq_id 1569409
- Location of start within SEQ ID NO 1246: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase family
- Location within SEQ ID NO 1248: from 7 to 254 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 586
- gi No. 4538897
- Description: (AL049482) AX110P-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 362
- Location of Alignment in SEQ ID NO 1248: from 1 to 362

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1249
- Ceres seq_id 1569410
- Location of start within SEQ ID NO 1246: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase family
- Location within SEQ ID NO 1249: from 1 to 243 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 587
- gi No. 4538897
- Description: (AL049482) AX110P-like protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 362
- Location of Alignment in SEQ ID NO 1249: from 1 to 351

Maximum Length Sequence:
related to:
Clone IDs:
25694
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1250
- Ceres seq_id 1569411
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1251
- Ceres seq_id 1569412
- Location of start within SEQ ID NO 1250: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 1251: from 166 to 254 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1252
- Ceres seq_id 1569413
- Location of start within SEQ ID NO 1250: at 234 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 1252: from 161 to 249 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1253
- Ceres seq_id 1569414
- Location of start within SEQ ID NO 1250: at 243 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 1253: from 158 to 246 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34420
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1254
- Ceres seq_id 1569427
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1255
- Ceres seq_id 1569428

- Location of start within SEQ ID NO 1254: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isocitrate and isopropylmalate dehydrogenases
- Location within SEQ ID NO 1255: from 51 to 399 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 588
- gi No. 4512615
- Description: (AC004793) Strong similarity to gb|X59970 3-isopropylmalate dehydrogenase (IMDH) from Brassica napus. EST gb|F14478 comes from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1255: from 6 to 409

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1256
- Ceres seq_id 1569429
- Location of start within SEQ ID NO 1254: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isocitrate and isopropylmalate dehydrogenases
- Location within SEQ ID NO 1256: from 46 to 394 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 589
- gi No. 4512615
- Description: (AC004793) Strong similarity to gb|X59970 3-isopropylmalate dehydrogenase (IMDH) from Brassica napus. EST gb|F14478 comes from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1256: from 1 to 404

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1257
- Ceres seq_id 1569430
- Location of start within SEQ ID NO 1254: at 260 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isocitrate and isopropylmalate dehydrogenases
- Location within SEQ ID NO 1257: from 1 to 313 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 590
- gi No. 4512615
- Description: (AC004793) Strong similarity to gb|X59970 3-isopropylmalate dehydrogenase (IMDH) from Brassica napus. EST gb|F14478 comes from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1257: from 1 to 323

Maximum Length Sequence:
related to:
Clone IDs:
36085

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1258
- Ceres seq_id 1569435

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1259
- Ceres seq_id 1569436
- Location of start within SEQ ID NO 1258: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1259: from 197 to 385 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 591
- gi No. 1709800
- Description: 26S PROTEASE REGULATORY SUBUNIT 8 (MSUG1 PROTEIN) (TAT-BINDING PROTEIN HOMOLOG 10) (TBP10) (P45/SUG) >gi|2137798|pir||S61923 SUG1 protein - mouse >gi|1165125|emb|CAA90961.1| (Z54219) mSUG1 protein [Mus musculus]
- % Identity: 82.7
- Alignment Length: 393
- Location of Alignment in SEQ ID NO 1259: from 27 to 419

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1260
- Ceres seq_id 1569437
- Location of start within SEQ ID NO 1258: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1260: from 183 to 371 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 592
- gi No. 1709800
- Description: 26S PROTEASE REGULATORY SUBUNIT 8 (MSUG1 PROTEIN) (TAT-BINDING PROTEIN HOMOLOG 10) (TBP10) (P45/SUG) >gi|2137798|pir||S61923 SUG1 protein - mouse >gi|1165125|emb|CAA90961.1| (Z54219) mSUG1 protein [Mus musculus]
- % Identity: 82.7
- Alignment Length: 393
- Location of Alignment in SEQ ID NO 1260: from 13 to 405

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1261
- Ceres seq_id 1569438
- Location of start within SEQ ID NO 1258: at 282 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 1261: from 128 to 316 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 593
- gi No. 1709800
- Description: 26S PROTEASE REGULATORY SUBUNIT 8 (MSUG1 PROTEIN) (TAT-BINDING PROTEIN HOMOLOG 10) (TBP10) (P45/SUG) >gi|2137798|pir||S61923

SUG1 protein - mouse >gi|1165125|emb|CAA90961.1| (Z54219) mSUG1 protein [Mus musculus]
- % Identity: 82.7
- Alignment Length: 393
- Location of Alignment in SEQ ID NO 1261: from 1 to 350

Maximum Length Sequence:
related to:
Clone IDs:
23909
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1262
- Ceres seq_id 1569439
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1263
- Ceres seq_id 1569440
- Location of start within SEQ ID NO 1262: at 450 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 594
- gi No. 882341
- Description: (U24702) LRP1 [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 218
- Location of Alignment in SEQ ID NO 1263: from 1 to 218

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1264
- Ceres seq_id 1569441
- Location of start within SEQ ID NO 1262: at 456 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 595
- gi No. 882341
- Description: (U24702) LRP1 [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 218
- Location of Alignment in SEQ ID NO 1264: from 1 to 216

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1265
- Ceres seq_id 1569442
- Location of start within SEQ ID NO 1262: at 537 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 596
- gi No. 882341
- Description: (U24702) LRP1 [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 218
- Location of Alignment in SEQ ID NO 1265: from 1 to 189

Maximum Length Sequence:
related to:
Clone IDs:

19562
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1266
- Ceres seq_id 1569451
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1267
- Ceres seq_id 1569452
- Location of start within SEQ ID NO 1266: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 597
- gi No. 4490749
- Description: (AL035708) 2-dehydro-3-deoxyphosphoheptonate aldolase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 1267: from 1 to 318

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1268
- Ceres seq_id 1569453
- Location of start within SEQ ID NO 1266: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 598
- gi No. 4490749
- Description: (AL035708) 2-dehydro-3-deoxyphosphoheptonate aldolase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 1268: from 1 to 302

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1269
- Ceres seq_id 1569454
- Location of start within SEQ ID NO 1266: at 228 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 599
- gi No. 4490749
- Description: (AL035708) 2-dehydro-3-deoxyphosphoheptonate aldolase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 318
- Location of Alignment in SEQ ID NO 1269: from 1 to 243

Maximum Length Sequence:
related to:
Clone IDs:
39734
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1270
- Ceres seq_id 1569470
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1271

- Ceres seq_id 1569471
- Location of start within SEQ ID NO 1270: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1271: from 140 to 370 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1272
- Ceres seq_id 1569472
- Location of start within SEQ ID NO 1270: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1272: from 75 to 305 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1273
- Ceres seq_id 1569473
- Location of start within SEQ ID NO 1270: at 291 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1273: from 44 to 274 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
150937
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1274
- Ceres seq_id 1569509

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1275
- Ceres seq_id 1569510
- Location of start within SEQ ID NO 1274: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1276
- Ceres seq_id 1569511
- Location of start within SEQ ID NO 1274: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1276: from 28 to 93 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1277
- Ceres seq_id 1569512
- Location of start within SEQ ID NO 1274: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
23238
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1278
- Ceres seq_id 1569522
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1279
- Ceres seq_id 1569523
- Location of start within SEQ ID NO 1278: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 600
- gi No. 3335374
- Description: (AC003028) glutaredoxin-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 1279: from 8 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1280
- Ceres seq_id 1569524
- Location of start within SEQ ID NO 1278: at 23 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 601
- gi No. 3335374
- Description: (AC003028) glutaredoxin-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 1280: from 1 to 176

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1281
- Ceres seq_id 1569525
- Location of start within SEQ ID NO 1278: at 580 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 602
- gi No. 3335374
- Description: (AC003028) glutaredoxin-like protein [Arabidopsis thaliana]

- % Identity: 99.1
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 1281: from 3 to 108

Maximum Length Sequence:
related to:
Clone IDs:
31388
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1282
- Ceres seq_id 1569526
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1283
- Ceres seq_id 1569527
- Location of start within SEQ ID NO 1282: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 603
- gi No. 710626
- Description: (D30719) ERD15 protein [Arabidopsis thaliana] >gi|3241941 (AC004625) dehydration-induced protein ERD15 [Arabidopsis thaliana] >gi|3894181 (AC005662) ERD15 protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 1283: from 1 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1284
- Ceres seq_id 1569528
- Location of start within SEQ ID NO 1282: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 604
- gi No. 710626
- Description: (D30719) ERD15 protein [Arabidopsis thaliana] >gi|3241941 (AC004625) dehydration-induced protein ERD15 [Arabidopsis thaliana] >gi|3894181 (AC005662) ERD15 protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 1284: from 1 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1285
- Ceres seq_id 1569529
- Location of start within SEQ ID NO 1282: at 312 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37122
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1286
- Ceres seq_id 1569530

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1287
- Ceres seq_id 1569531
- Location of start within SEQ ID NO 1286: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 605
- gi No. 5441877
- Description: (AP000367) Similar to glycogenin glucosyltransferase (EC 2.4.1.186). (Z97341) [Oryza sativa]
- % Identity: 73.6
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1287: from 1 to 273

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1288
- Ceres seq_id 1569532
- Location of start within SEQ ID NO 1286: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 606
- gi No. 5441877
- Description: (AP000367) Similar to glycogenin glucosyltransferase (EC 2.4.1.186). (Z97341) [Oryza sativa]
- % Identity: 73.6
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1288: from 1 to 244

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1289
- Ceres seq_id 1569533
- Location of start within SEQ ID NO 1286: at 240 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 607
- gi No. 5441877
- Description: (AP000367) Similar to glycogenin glucosyltransferase (EC 2.4.1.186). (Z97341) [Oryza sativa]
- % Identity: 73.6
- Alignment Length: 275
- Location of Alignment in SEQ ID NO 1289: from 1 to 194

Maximum Length Sequence:
related to:
Clone IDs:
39204
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1290
- Ceres seq_id 1569538

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1291
- Ceres seq_id 1569539
- Location of start within SEQ ID NO 1290: at 162 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1292
- Ceres seq_id 1569540
- Location of start within SEQ ID NO 1290: at 363 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1293
- Ceres seq_id 1569541
- Location of start within SEQ ID NO 1290: at 817 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1293: from 41 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 608
- gi No. 2864602
- Description: (Y12071) thylakoid lumen rotamase [Spinacia oleracea]
- % Identity: 83.6
- Alignment Length: 55
- Location of Alignment in SEQ ID NO 1293: from 14 to 68

Maximum Length Sequence:
related to:
Clone IDs:
234
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1294
- Ceres seq_id 1569546
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1295
- Ceres seq_id 1569547
- Location of start within SEQ ID NO 1294: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1295: from 29 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 609
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 91.9
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 1295: from 23 to 153

Maximum Length Sequence:

related to:
Clone IDs:
9338
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1296
- Ceres seq_id 1569552
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1297
- Ceres seq_id 1569553
- Location of start within SEQ ID NO 1296: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 1297: from 25 to 140 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1298
- Ceres seq_id 1569554
- Location of start within SEQ ID NO 1296: at 190 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 1298: from 1 to 116 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1299
- Ceres seq_id 1569555
- Location of start within SEQ ID NO 1296: at 199 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 1299: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32727
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1300
- Ceres seq_id 1569579
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1301
- Ceres seq_id 1569580
- Location of start within SEQ ID NO 1300: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 610
- gi No. 2618721
- Description: (U49072) IAA16 [Arabidopsis thaliana]
- % Identity: 100

- Alignment Length: 236
- Location of Alignment in SEQ ID NO 1301: from 63 to 298

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1302
- Ceres seq_id 1569581
- Location of start within SEQ ID NO 1300: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 611
- gi No. 2618721
- Description: (U49072) IAA16 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 236
- Location of Alignment in SEQ ID NO 1302: from 1 to 236

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1303
- Ceres seq_id 1569582
- Location of start within SEQ ID NO 1300: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 612
- gi No. 2618721
- Description: (U49072) IAA16 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 236
- Location of Alignment in SEQ ID NO 1303: from 1 to 215

Maximum Length Sequence:
related to:
Clone IDs:
24400
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1304
- Ceres seq_id 1569583
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1305
- Ceres seq_id 1569584
- Location of start within SEQ ID NO 1304: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1305: from 84 to 194 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 613
- gi No. 4982517
- Description: (AC000107) F17F8.23 [Arabidopsis thaliana]
- % Identity: 79.1
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 1305: from 104 to 194

Maximum Length Sequence:
related to:
Clone IDs:

36468
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1306
- Ceres seq_id 1569585
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1307
- Ceres seq_id 1569586
- Location of start within SEQ ID NO 1306: at 558 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 614
- gi No. 4262179
- Description: (AC005508) 27668 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1307: from 1 to 201

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1308
- Ceres seq_id 1569587
- Location of start within SEQ ID NO 1306: at 609 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 615
- gi No. 4262179
- Description: (AC005508) 27668 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1308: from 1 to 184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1309
- Ceres seq_id 1569588
- Location of start within SEQ ID NO 1306: at 642 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 616
- gi No. 4262179
- Description: (AC005508) 27668 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1309: from 1 to 173

Maximum Length Sequence:
related to:
Clone IDs:
151415
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1310
- Ceres seq_id 1569589
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1311
- Ceres seq_id 1569590
- Location of start within SEQ ID NO 1310: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoprotein amino terminal region
- Location within SEQ ID NO 1311: from 130 to 321 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1312
- Ceres seq_id 1569591
- Location of start within SEQ ID NO 1310: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoprotein amino terminal region
- Location within SEQ ID NO 1312: from 94 to 285 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1313
- Ceres seq_id 1569592
- Location of start within SEQ ID NO 1310: at 333 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoprotein amino terminal region
- Location within SEQ ID NO 1313: from 45 to 236 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
151603
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1314
- Ceres seq_id 1569603

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1315
- Ceres seq_id 1569604
- Location of start within SEQ ID NO 1314: at 216 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 1315: from 62 to 278 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 617
- gi No. 4678353
- Description: (AL049659) cysteine endopeptidase precursor-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 276
- Location of Alignment in SEQ ID NO 1315: from 25 to 300

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1316
- Ceres seq_id 1569605
- Location of start within SEQ ID NO 1314: at 225 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 1316: from 59 to 275 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 618
- gi No. 4678353
- Description: (AL049659) cysteine endopeptidase precursor-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 276
- Location of Alignment in SEQ ID NO 1316: from 22 to 297

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1317
- Ceres seq_id 1569606
- Location of start within SEQ ID NO 1314: at 342 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 1317: from 20 to 236 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 619
- gi No. 4678353
- Description: (AL049659) cysteine endopeptidase precursor-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 276
- Location of Alignment in SEQ ID NO 1317: from 1 to 258

Maximum Length Sequence:
related to:
Clone IDs:
151953
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1318
- Ceres seq_id 1569611

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1319
- Ceres seq_id 1569612
- Location of start within SEQ ID NO 1318: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter
- Location within SEQ ID NO 1319: from 79 to 265 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1320
- Ceres seq_id 1569613
- Location of start within SEQ ID NO 1318: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ABC transporter

- Location within SEQ ID NO 1320: from 74 to 260 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
152227
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1321
- Ceres seq_id 1569614
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1322
- Ceres seq_id 1569615
- Location of start within SEQ ID NO 1321: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1323
- Ceres seq_id 1569616
- Location of start within SEQ ID NO 1321: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1324
- Ceres seq_id 1569617
- Location of start within SEQ ID NO 1321: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 620
- gi No. 1707008
- Description: (U78721) 30S ribosomal protein S5 isolog [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 1324: from 1 to 117

Maximum Length Sequence:
related to:
Clone IDs:
152232
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1325
- Ceres seq_id 1569618
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1326
- Ceres seq_id 1569619
- Location of start within SEQ ID NO 1325: at 327 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1326: from 1 to 167 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1327
- Ceres seq_id 1569620
- Location of start within SEQ ID NO 1325: at 459 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1327: from 1 to 123 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1328
- Ceres seq_id 1569621
- Location of start within SEQ ID NO 1325: at 507 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0023
- Location within SEQ ID NO 1328: from 1 to 107 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
35056
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1329
- Ceres seq_id 1569652

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1330
- Ceres seq_id 1569653
- Location of start within SEQ ID NO 1329: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 621
- gi No. 2088653
- Description: (AF002109) Hs1pro-1 related protein isolog [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 435
- Location of Alignment in SEQ ID NO 1330: from 24 to 458

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1331
- Ceres seq_id 1569654
- Location of start within SEQ ID NO 1329: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 622
- gi No. 2088653

- Description: (AF002109) Hs1pro-1 related protein isolog [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 435
- Location of Alignment in SEQ ID NO 1331: from 1 to 435

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1332
- Ceres seq_id 1569655
- Location of start within SEQ ID NO 1329: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 623
- gi No. 2088653
- Description: (AF002109) Hs1pro-1 related protein isolog [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 435
- Location of Alignment in SEQ ID NO 1332: from 1 to 432

Maximum Length Sequence:
related to:
Clone IDs:
32891
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1333
- Ceres seq_id 1569656
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1334
- Ceres seq_id 1569657
- Location of start within SEQ ID NO 1333: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1334: from 93 to 131 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 624
- gi No. 2494896
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 3 DELTA SUBUNIT (EIF-3 DELTA) (EIF3 P36) (TGF-BETA RECEPTOR INTERACTING PROTEIN 1) (TRIP-1) >gi|2129749|pir||S60256 TGF-beta receptor interacting protein 1 homolog - Arabidopsis thaliana
- % Identity: 88.4
- Alignment Length: 328
- Location of Alignment in SEQ ID NO 1334: from 52 to 379

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1335
- Ceres seq_id 1569658
- Location of start within SEQ ID NO 1333: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1335: from 69 to 107 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 625
- gi No. 2494896
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 3 DELTA SUBUNIT (EIF-3 DELTA) (EIF3 P36) (TGF-BETA RECEPTOR INTERACTING PROTEIN 1) (TRIP-1) >gi|2129749|pir||S60256 TGF-beta receptor interacting protein 1 homolog - Arabidopsis thaliana
- % Identity: 88.4
- Alignment Length: 328
- Location of Alignment in SEQ ID NO 1335: from 28 to 355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1336
- Ceres seq_id 1569659
- Location of start within SEQ ID NO 1333: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1336: from 54 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 626
- gi No. 2494896
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 3 DELTA SUBUNIT (EIF-3 DELTA) (EIF3 P36) (TGF-BETA RECEPTOR INTERACTING PROTEIN 1) (TRIP-1) >gi|2129749|pir||S60256 TGF-beta receptor interacting protein 1 homolog - Arabidopsis thaliana
- % Identity: 88.4
- Alignment Length: 328
- Location of Alignment in SEQ ID NO 1336: from 13 to 340

Maximum Length Sequence:
related to:
Clone IDs:
36412
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1337
- Ceres seq_id 1569667

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1338
- Ceres seq_id 1569668
- Location of start within SEQ ID NO 1337: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1338: from 310 to 344 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 627
- gi No. 3152583
- Description: (AC002986) Contains similarity to inhibitor of apoptosis protein gb|U45881 from D. melanogaster. [Arabidopsis thaliana]
- % Identity: 95.4
- Alignment Length: 347
- Location of Alignment in SEQ ID NO 1338: from 23 to 358

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1339
- Ceres seq_id 1569669
- Location of start within SEQ ID NO 1337: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1339: from 288 to 322 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 628
- gi No. 3152583
- Description: (AC002986) Contains similarity to inhibitor of apoptosis protein gb|U45881 from D. melanogaster. [Arabidopsis thaliana]
- % Identity: 95.4
- Alignment Length: 347
- Location of Alignment in SEQ ID NO 1339: from 1 to 336

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1340
- Ceres seq_id 1569670
- Location of start within SEQ ID NO 1337: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1340: from 283 to 317 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 629
- gi No. 3152583
- Description: (AC002986) Contains similarity to inhibitor of apoptosis protein gb|U45881 from D. melanogaster. [Arabidopsis thaliana]
- % Identity: 95.4
- Alignment Length: 347
- Location of Alignment in SEQ ID NO 1340: from 1 to 331

Maximum Length Sequence:
related to:
Clone IDs:
35974
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1341
- Ceres seq_id 1569671

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1342
- Ceres seq_id 1569672
- Location of start within SEQ ID NO 1341: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1342: from 49 to 133 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1343
- Ceres seq_id 1569673
- Location of start within SEQ ID NO 1341: at 489 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase

- Location within SEQ ID NO 1343: from 1 to 122 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1344
- Ceres seq_id 1569674
- Location of start within SEQ ID NO 1341: at 501 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1344: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
36399
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1345
- Ceres seq_id 1569685
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1346
- Ceres seq_id 1569686
- Location of start within SEQ ID NO 1345: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1346: from 240 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 630
- gi No. 4850408
- Description: (AC007357) Contains PF|00097 Zinc finger (C3HC4) ring finger motif. [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 258
- Location of Alignment in SEQ ID NO 1346: from 81 to 338

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1347
- Ceres seq_id 1569687
- Location of start within SEQ ID NO 1345: at 351 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1347: from 124 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 631
- gi No. 4850408
- Description: (AC007357) Contains PF|00097 Zinc finger (C3HC4) ring finger motif. [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 258
- Location of Alignment in SEQ ID NO 1347: from 1 to 222

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1348
- Ceres seq_id 1569688
- Location of start within SEQ ID NO 1345: at 378 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1348: from 115 to 152 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 632
- gi No. 4850408
- Description: (AC007357) Contains PF|00097 Zinc finger (C3HC4) ring finger motif. [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 258
- Location of Alignment in SEQ ID NO 1348: from 1 to 213

Maximum Length Sequence:
related to:
Clone IDs:
153949
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1349
- Ceres seq_id 1569697
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1350
- Ceres seq_id 1569698
- Location of start within SEQ ID NO 1349: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 1350: from 19 to 107 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1351
- Ceres seq_id 1569699
- Location of start within SEQ ID NO 1349: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 1351: from 1 to 89 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1352
- Ceres seq_id 1569700
- Location of start within SEQ ID NO 1349: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 1352: from 1 to 84 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
154375
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1353
- Ceres seq_id 1569717
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1354
- Ceres seq_id 1569718
- Location of start within SEQ ID NO 1353: at 26 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl transferases group 1
- Location within SEQ ID NO 1354: from 200 to 371 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 633
- gi No. 3834314
- Description: (AC005679) Similar to gene pi010 glycosyltransferase gi|2257490 from S. pombe clone 1750 gb|AB004534. ESTs gb|T46079 and gb|AA394466 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1354: from 1 to 403

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1355
- Ceres seq_id 1569719
- Location of start within SEQ ID NO 1353: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl transferases group 1
- Location within SEQ ID NO 1355: from 192 to 363 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 634
- gi No. 3834314
- Description: (AC005679) Similar to gene pi010 glycosyltransferase gi|2257490 from S. pombe clone 1750 gb|AB004534. ESTs gb|T46079 and gb|AA394466 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1355: from 1 to 395

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1356
- Ceres seq_id 1569720
- Location of start within SEQ ID NO 1353: at 461 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl transferases group 1
- Location within SEQ ID NO 1356: from 55 to 226 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 635
- gi No. 3834314

- Description: (AC005679) Similar to gene pi010 glycosyltransferase gi|2257490 from S. pombe clone 1750 gb|AB004534. ESTs gb|T46079 and gb|AA394466 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1356: from 1 to 258

Maximum Length Sequence:
related to:
Clone IDs:
154387
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1357
- Ceres seq_id 1569721
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1358
- Ceres seq_id 1569722
- Location of start within SEQ ID NO 1357: at 55 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 1358: from 85 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 636
- gi No. 1418329
- Description: (X98677) zinc finger protein [Arabidopsis thaliana] >gi|1418341|emb|CAA67234| (X98676) zinc finger protein [Arabidopsis thaliana]
- % Identity: 87.2
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 1358: from 1 to 117

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1359
- Ceres seq_id 1569723
- Location of start within SEQ ID NO 1357: at 112 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 1359: from 66 to 88 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 637
- gi No. 1418329
- Description: (X98677) zinc finger protein [Arabidopsis thaliana] >gi|1418341|emb|CAA67234| (X98676) zinc finger protein [Arabidopsis thaliana]
- % Identity: 87.2
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 1359: from 1 to 98

Maximum Length Sequence:
related to:
Clone IDs:
154741
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1360
- Ceres seq_id 1569724
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1361

- Ceres seq_id 1569725
- Location of start within SEQ ID NO 1360: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 1361: from 90 to 332 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 638
- gi No. 3193316
- Description: (AF069299) contains similarity to nucleotide sugar epimerases [Arabidopsis thaliana]
- % Identity: 70.4
- Alignment Length: 340
- Location of Alignment in SEQ ID NO 1361: from 1 to 332

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1362
- Ceres seq_id 1569726
- Location of start within SEQ ID NO 1360: at 258 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 1362: from 56 to 298 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 639
- gi No. 3193316
- Description: (AF069299) contains similarity to nucleotide sugar epimerases [Arabidopsis thaliana]
- % Identity: 70.4
- Alignment Length: 340
- Location of Alignment in SEQ ID NO 1362: from 1 to 298

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1363
- Ceres seq_id 1569727
- Location of start within SEQ ID NO 1360: at 654 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- NAD dependent epimerase/dehydratase family
- Location within SEQ ID NO 1363: from 1 to 166 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 640
- gi No. 3193316
- Description: (AF069299) contains similarity to nucleotide sugar epimerases [Arabidopsis thaliana]
- % Identity: 70.4
- Alignment Length: 340
- Location of Alignment in SEQ ID NO 1363: from 1 to 166

Maximum Length Sequence:
related to:
Clone IDs:
155041
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1364

- Ceres seq_id 1569732

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1365
- Ceres seq_id 1569733
- Location of start within SEQ ID NO 1364: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1365: from 92 to 368 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1366
- Ceres seq_id 1569734
- Location of start within SEQ ID NO 1364: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1366: from 63 to 339 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1367
- Ceres seq_id 1569735
- Location of start within SEQ ID NO 1364: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1367: from 45 to 321 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3434
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1368
- Ceres seq_id 1569744

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1369
- Ceres seq_id 1569745
- Location of start within SEQ ID NO 1368: at 420 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1369: from 139 to 189 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 641
- gi No. 3176664
- Description: (AC004393) Contains similarity to beta scruin gb|Z47541 from Limulus polyphemus. ESTs gb|T04493 and gb|AA585955 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.8

- Alignment Length: 227
- Location of Alignment in SEQ ID NO 1369: from 1 to 211

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1370
- Ceres seq_id 1569746
- Location of start within SEQ ID NO 1368: at 468 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1370: from 123 to 173 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 642
- gi No. 3176664
- Description: (AC004393) Contains similarity to beta scruin gb|Z47541 from Limulus polyphemus. ESTs gb|T04493 and gb|AA585955 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.8
- Alignment Length: 227
- Location of Alignment in SEQ ID NO 1370: from 1 to 195

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1371
- Ceres seq_id 1569747
- Location of start within SEQ ID NO 1368: at 522 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1371: from 105 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 643
- gi No. 3176664
- Description: (AC004393) Contains similarity to beta scruin gb|Z47541 from Limulus polyphemus. ESTs gb|T04493 and gb|AA585955 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.8
- Alignment Length: 227
- Location of Alignment in SEQ ID NO 1371: from 1 to 177

Maximum Length Sequence:
related to:
Clone IDs:
3016
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1372
- Ceres seq_id 1569748

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1373
- Ceres seq_id 1569749
- Location of start within SEQ ID NO 1372: at 750 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase molybdopterin binding domain
- Location within SEQ ID NO 1373: from 1 to 147 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1374
- Ceres seq_id 1569750
- Location of start within SEQ ID NO 1372: at 768 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase molybdopterin binding domain
- Location within SEQ ID NO 1374: from 1 to 141 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1375
- Ceres seq_id 1569751
- Location of start within SEQ ID NO 1372: at 825 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase molybdopterin binding domain
- Location within SEQ ID NO 1375: from 1 to 122 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
2450
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1376
- Ceres seq_id 1569760
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1377
- Ceres seq_id 1569761
- Location of start within SEQ ID NO 1376: at 547 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 644
- gi No. 3212877
- Description: (AC004005) Lea-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 183
- Location of Alignment in SEQ ID NO 1377: from 1 to 169

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1378
- Ceres seq_id 1569762
- Location of start within SEQ ID NO 1376: at 739 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 645
- gi No. 3212877
- Description: (AC004005) Lea-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 183
- Location of Alignment in SEQ ID NO 1378: from 1 to 105

Maximum Length Sequence:
related to:
Clone IDs:
37980
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1379
- Ceres seq_id 1569775
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1380
- Ceres seq_id 1569776
- Location of start within SEQ ID NO 1379: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 'chromo' (CHRromatin Organization MOdifier) domain
- Location within SEQ ID NO 1380: from 330 to 369 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 646
- gi No. 4102582
- Description: (AF013115) CAO [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 1380: from 12 to 385

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1381
- Ceres seq_id 1569777
- Location of start within SEQ ID NO 1379: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 'chromo' (CHRromatin Organization MOdifier) domain
- Location within SEQ ID NO 1381: from 319 to 358 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 647
- gi No. 4102582
- Description: (AF013115) CAO [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 1381: from 1 to 374

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1382
- Ceres seq_id 1569778
- Location of start within SEQ ID NO 1379: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 'chromo' (CHRromatin Organization MOdifier) domain
- Location within SEQ ID NO 1382: from 312 to 351 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 648
- gi No. 4102582
- Description: (AF013115) CAO [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 1382: from 1 to 367

Maximum Length Sequence:
related to:
Clone IDs:
33380
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1383
- Ceres seq_id 1569787
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1384
- Ceres seq_id 1569788
- Location of start within SEQ ID NO 1383: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Assemblin (Peptidase family S21)
- Location within SEQ ID NO 1384: from 129 to 205 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1385
- Ceres seq_id 1569789
- Location of start within SEQ ID NO 1383: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Assemblin (Peptidase family S21)
- Location within SEQ ID NO 1385: from 124 to 200 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1386
- Ceres seq_id 1569790
- Location of start within SEQ ID NO 1383: at 787 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Assemblin (Peptidase family S21)
- Location within SEQ ID NO 1386: from 1 to 63 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
25785
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1387
- Ceres seq_id 1569799
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1388
- Ceres seq_id 1569800
- Location of start within SEQ ID NO 1387: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Actin
- Location within SEQ ID NO 1388: from 5 to 377 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 649
- gi No. 1703115
- Description: ACTIN 3 >gi|2129526|pir||S68112 actin 3 - Arabidopsis thaliana >gi|1145695 (U39480) actin [Arabidopsis thaliana] >gi|3236244 (AC004684) actin 3 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 377
- Location of Alignment in SEQ ID NO 1388: from 1 to 377

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1389
- Ceres seq_id 1569801
- Location of start within SEQ ID NO 1387: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Actin
- Location within SEQ ID NO 1389: from 1 to 360 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 650
- gi No. 1703115
- Description: ACTIN 3 >gi|2129526|pir||S68112 actin 3 - Arabidopsis thaliana >gi|1145695 (U39480) actin [Arabidopsis thaliana] >gi|3236244 (AC004684) actin 3 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 377
- Location of Alignment in SEQ ID NO 1389: from 1 to 360

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1390
- Ceres seq_id 1569802
- Location of start within SEQ ID NO 1387: at 243 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Actin
- Location within SEQ ID NO 1390: from 1 to 332 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 651
- gi No. 1703115
- Description: ACTIN 3 >gi|2129526|pir||S68112 actin 3 - Arabidopsis thaliana >gi|1145695 (U39480) actin [Arabidopsis thaliana] >gi|3236244 (AC004684) actin 3 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 377
- Location of Alignment in SEQ ID NO 1390: from 1 to 332

Maximum Length Sequence:
related to:
Clone IDs:
31774
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1391
- Ceres seq_id 1569807
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1392
- Ceres seq_id 1569808
- Location of start within SEQ ID NO 1391: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1392: from 333 to 370 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 652
- gi No. 4337011
- Description: (AF119572) zinc-binding peroxisomal integral membrane protein [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 381
- Location of Alignment in SEQ ID NO 1392: from 7 to 387

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1393
- Ceres seq_id 1569809
- Location of start within SEQ ID NO 1391: at 19 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1393: from 327 to 364 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 653
- gi No. 4337011
- Description: (AF119572) zinc-binding peroxisomal integral membrane protein [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 381
- Location of Alignment in SEQ ID NO 1393: from 1 to 381

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1394
- Ceres seq_id 1569810
- Location of start within SEQ ID NO 1391: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1394: from 293 to 330 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 654
- gi No. 4337011
- Description: (AF119572) zinc-binding peroxisomal integral membrane protein [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 381
- Location of Alignment in SEQ ID NO 1394: from 1 to 347

Maximum Length Sequence:
related to:
Clone IDs:
4598
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1395
- Ceres seq_id 1569815
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1396
- Ceres seq_id 1569816
- Location of start within SEQ ID NO 1395: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- eubacterial secY protein
- Location within SEQ ID NO 1396: from 77 to 460 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 655
- gi No. 3834321
- Description: (AC005679) Strong similarity to F13P17.9 gi|3337356 transport protein SEC61 alpha subunit homolog from Arabidopsis thaliana BAC gb|AC004481. [Arabidopsis thaliana]
- % Identity: 89.7
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 1396: from 1 to 474

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1397
- Ceres seq_id 1569817
- Location of start within SEQ ID NO 1395: at 363 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- eubacterial secY protein
- Location within SEQ ID NO 1397: from 11 to 394 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 656
- gi No. 3834321
- Description: (AC005679) Strong similarity to F13P17.9 gi|3337356 transport protein SEC61 alpha subunit homolog from Arabidopsis thaliana BAC gb|AC004481. [Arabidopsis thaliana]
- % Identity: 89.7
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 1397: from 1 to 408

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1398
- Ceres seq_id 1569818
- Location of start within SEQ ID NO 1395: at 399 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- eubacterial secY protein
- Location within SEQ ID NO 1398: from 1 to 382 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 657
- gi No. 3834321
- Description: (AC005679) Strong similarity to F13P17.9 gi|3337356 transport protein SEC61 alpha subunit homolog from Arabidopsis thaliana BAC gb|AC004481. [Arabidopsis thaliana]
- % Identity: 89.7
- Alignment Length: 474
- Location of Alignment in SEQ ID NO 1398: from 1 to 396

Maximum Length Sequence:
related to:

Clone IDs:
29991
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1399
- Ceres seq_id 1569834
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1400
- Ceres seq_id 1569835
- Location of start within SEQ ID NO 1399: at 245 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1400: from 7 to 405 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 658
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 1400: from 1 to 449

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1401
- Ceres seq_id 1569836
- Location of start within SEQ ID NO 1399: at 389 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1401: from 1 to 357 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 659
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 1401: from 1 to 401

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1402
- Ceres seq_id 1569837
- Location of start within SEQ ID NO 1399: at 548 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 1402: from 1 to 304 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 660
- gi No. 119143

- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 1402: from 1 to 348

Maximum Length Sequence:
related to:
Clone IDs:
155616
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1403
- Ceres seq_id 1569846

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1404
- Ceres seq_id 1569847
- Location of start within SEQ ID NO 1403: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Intermediate filament proteins
- Location within SEQ ID NO 1404: from 8 to 81 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1405
- Ceres seq_id 1569848
- Location of start within SEQ ID NO 1403: at 22 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Intermediate filament proteins
- Location within SEQ ID NO 1405: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1406
- Ceres seq_id 1569849
- Location of start within SEQ ID NO 1403: at 46 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Intermediate filament proteins
- Location within SEQ ID NO 1406: from 1 to 66 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
156172
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1407
- Ceres seq_id 1569867

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1408
- Ceres seq_id 1569868

- Location of start within SEQ ID NO 1407: at 428 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1408: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 661
- gi No. 1532165
- Description: (U63815) similar to dehydrogenase encoded by GenBank Accession Number S39508; localized according to blastn similarity to EST sequences; therefore, the coding span corresponds only to an area
- % Identity: 99.2
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 1408: from 4 to 135

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1409
- Ceres seq_id 1569869
- Location of start within SEQ ID NO 1407: at 458 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1409: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 662
- gi No. 1532165
- Description: (U63815) similar to dehydrogenase encoded by GenBank Accession Number S39508; localized according to blastn similarity to EST sequences; therefore, the coding span corresponds only to an area
- % Identity: 99.2
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 1409: from 1 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1410
- Ceres seq_id 1569870
- Location of start within SEQ ID NO 1407: at 494 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 663
- gi No. 1532165
- Description: (U63815) similar to dehydrogenase encoded by GenBank Accession Number S39508; localized according to blastn similarity to EST sequences; therefore, the coding span corresponds only to an area
- % Identity: 99.2
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 1410: from 1 to 113

Maximum Length Sequence:
related to:
Clone IDs:
156484
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1411
- Ceres seq_id 1569871

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1412
- Ceres seq_id 1569872
- Location of start within SEQ ID NO 1411: at 414 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 664
- gi No. 2244999
- Description: (Z97341) transcription factor like protein [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 1412: from 373 to 385

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1413
- Ceres seq_id 1569873
- Location of start within SEQ ID NO 1411: at 420 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 665
- gi No. 2244999
- Description: (Z97341) transcription factor like protein [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 1413: from 371 to 383

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1414
- Ceres seq_id 1569874
- Location of start within SEQ ID NO 1411: at 537 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 666
- gi No. 2244999
- Description: (Z97341) transcription factor like protein [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 1414: from 332 to 344

Maximum Length Sequence:
related to:
Clone IDs:
156940
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1415
- Ceres seq_id 1569897
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1416
- Ceres seq_id 1569898
- Location of start within SEQ ID NO 1415: at 189 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1416: from 181 to 221 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1417
- Ceres seq_id 1569899
- Location of start within SEQ ID NO 1415: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1417: from 165 to 205 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
158231
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1418
- Ceres seq_id 1569932
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1419
- Ceres seq_id 1569933
- Location of start within SEQ ID NO 1418: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1419: from 9 to 49 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1420
- Ceres seq_id 1569934
- Location of start within SEQ ID NO 1418: at 7 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1420: from 7 to 47 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1421
- Ceres seq_id 1569935
- Location of start within SEQ ID NO 1418: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
158515
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1422
- Ceres seq_id 1569942
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1423
- Ceres seq_id 1569943
- Location of start within SEQ ID NO 1422: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 667
- gi No. 5732040
- Description: (AF147262) contains similarity to mouse and human SL15 proteins (GB:AF038961 and U41996) [Arabidopsis thaliana]
- % Identity: 84.8
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 1423: from 1 to 105

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1424
- Ceres seq_id 1569945
- Location of start within SEQ ID NO 1422: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 668
- gi No. 5732040
- Description: (AF147262) contains similarity to mouse and human SL15 proteins (GB:AF038961 and U41996) [Arabidopsis thaliana]
- % Identity: 84.8
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 1424: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
158584
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1425
- Ceres seq_id 1569946
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1426
- Ceres seq_id 1569947
- Location of start within SEQ ID NO 1425: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 1426: from 76 to 147 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1427
- Ceres seq_id 1569948
- Location of start within SEQ ID NO 1425: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 1427: from 54 to 125 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1428
- Ceres seq_id 1569949
- Location of start within SEQ ID NO 1425: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mitochondrial carrier proteins
- Location within SEQ ID NO 1428: from 39 to 110 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
158702
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1429
- Ceres seq_id 1569950
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1430
- Ceres seq_id 1569951
- Location of start within SEQ ID NO 1429: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Inositol monophosphatase family
- Location within SEQ ID NO 1430: from 148 to 211 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 669
- gi No. 5732074
- Description: (AF162444) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 1430: from 1 to 67

Maximum Length Sequence:
related to:
Clone IDs:
159337
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1431
- Ceres seq_id 1569972
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1432
- Ceres seq_id 1569973
- Location of start within SEQ ID NO 1431: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1432: from 23 to 69 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1433
- Ceres seq_id 1569974
- Location of start within SEQ ID NO 1431: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1433: from 1 to 46 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1434
- Ceres seq_id 1569975
- Location of start within SEQ ID NO 1431: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
159568
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1435
- Ceres seq_id 1569986
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1436
- Ceres seq_id 1569987
- Location of start within SEQ ID NO 1435: at 305 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Tetrapyrrole (Corrin/Porphyrin) Methylases.
- Location within SEQ ID NO 1436: from 116 to 329 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 670
- gi No. 1146165
- Description: (L47479) urophorphyrin III methylase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 369
- Location of Alignment in SEQ ID NO 1436: from 1 to 369

Maximum Length Sequence:
related to:
Clone IDs:
40123
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1437
- Ceres seq_id 1570000
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1438
- Ceres seq_id 1570001
- Location of start within SEQ ID NO 1437: at 135 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 1438: from 61 to 310 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1439
- Ceres seq_id 1570002
- Location of start within SEQ ID NO 1437: at 387 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 1439: from 1 to 226 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1440
- Ceres seq_id 1570003
- Location of start within SEQ ID NO 1437: at 474 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 1440: from 1 to 197 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39002
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1441
- Ceres seq_id 1570004
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1442
- Ceres seq_id 1570005
- Location of start within SEQ ID NO 1441: at 446 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Tubulin
- Location within SEQ ID NO 1442: from 4 to 306 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 671
- gi No. 267073
- Description: TUBULIN BETA-2/BETA-3 CHAIN >gi|320184|pir||JQ1587 tubulin beta chain - Arabidopsis thaliana >gi|166898 (M84700) beta-2 tubulin [Arabidopsis thaliana] >gi|166900 (M84701) beta-3 tubulin [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 1442: from 4 to 333

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1443
- Ceres seq_id 1570006
- Location of start within SEQ ID NO 1441: at 533 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Tubulin
- Location within SEQ ID NO 1443: from 1 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 672
- gi No. 267073
- Description: TUBULIN BETA-2/BETA-3 CHAIN >gi|320184|pir||JQ1587 tubulin beta chain - Arabidopsis thaliana >gi|166898 (M84700) beta-2 tubulin [Arabidopsis thaliana] >gi|166900 (M84701) beta-3 tubulin [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 1443: from 1 to 304

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1444
- Ceres seq_id 1570007
- Location of start within SEQ ID NO 1441: at 581 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Tubulin
- Location within SEQ ID NO 1444: from 1 to 261 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 673
- gi No. 267073
- Description: TUBULIN BETA-2/BETA-3 CHAIN >gi|320184|pir||JQ1587 tubulin beta chain - Arabidopsis thaliana >gi|166898 (M84700) beta-2 tubulin [Arabidopsis thaliana] >gi|166900 (M84701) beta-3 tubulin [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 1444: from 1 to 288

Maximum Length Sequence:
related to:
Clone IDs:
17506
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1445
- Ceres seq_id 1570008

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1446
- Ceres seq_id 1570009
- Location of start within SEQ ID NO 1445: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 674
- gi No. 4584357
- Description: (AC006420) photosystem II protein X precursor [Arabidopsis thaliana]
- % Identity: 100

- Alignment Length: 116
- Location of Alignment in SEQ ID NO 1446: from 23 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1447
- Ceres seq_id 1570010
- Location of start within SEQ ID NO 1445: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 675
- gi No. 4584357
- Description: (AC006420) photosystem II protein X precursor [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 1447: from 1 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1448
- Ceres seq_id 1570011
- Location of start within SEQ ID NO 1445: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 676
- gi No. 4584357
- Description: (AC006420) photosystem II protein X precursor [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 1448: from 1 to 110

Maximum Length Sequence:
related to:
Clone IDs:
18250
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1449
- Ceres seq_id 1570036
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1450
- Ceres seq_id 1570037
- Location of start within SEQ ID NO 1449: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 677
- gi No. 5734712
- Description: (AC008075) ESTs gb|T41781 and gb|AA586096 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 1450: from 1 to 106

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1451
- Ceres seq_id 1570038

- Location of start within SEQ ID NO 1449: at 193 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 678
- gi No. 5734712
- Description: (AC008075) ESTs gb|T41781 and gb|AA586096 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 1451: from 1 to 68

Maximum Length Sequence:
related to:
Clone IDs:
18253

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1452
- Ceres seq_id 1570039

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1453
- Ceres seq_id 1570040
- Location of start within SEQ ID NO 1452: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 679
- gi No. 3193306
- Description: (AF069300) contains similarity to Arabidopsis membrane-associated salt-inducible-like protein (GB:AL021637) [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 210
- Location of Alignment in SEQ ID NO 1453: from 12 to 129

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1454
- Ceres seq_id 1570041
- Location of start within SEQ ID NO 1452: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 680
- gi No. 3193306
- Description: (AF069300) contains similarity to Arabidopsis membrane-associated salt-inducible-like protein (GB:AL021637) [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 210
- Location of Alignment in SEQ ID NO 1454: from 1 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1455
- Ceres seq_id 1570042
- Location of start within SEQ ID NO 1452: at 466 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 681
- gi No. 3193306
- Description: (AF069300) contains similarity to Arabidopsis membrane-associated salt-inducible-like protein (GB:AL021637) [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 210
- Location of Alignment in SEQ ID NO 1455: from 1 to 65

Maximum Length Sequence:
related to:
Clone IDs:
18662
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1456
- Ceres seq_id 1570066
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1457
- Ceres seq_id 1570067
- Location of start within SEQ ID NO 1456: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 682
- gi No. 4193382
- Description: (AF083336) ribosomal protein S27 [Arabidopsis thaliana] >gi|4193384 (AF083337) ribosomal protein S27 [Arabidopsis thaliana]
- % Identity: 94.2
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 1457: from 1 to 84

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1458
- Ceres seq_id 1570068
- Location of start within SEQ ID NO 1456: at 185 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 683
- gi No. 4193382
- Description: (AF083336) ribosomal protein S27 [Arabidopsis thaliana] >gi|4193384 (AF083337) ribosomal protein S27 [Arabidopsis thaliana]
- % Identity: 94.2
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 1458: from 1 to 50

Maximum Length Sequence:
related to:
Clone IDs:
19020
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1459
- Ceres seq_id 1570069
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1460
- Ceres seq_id 1570070
- Location of start within SEQ ID NO 1459: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1460: from 34 to 188 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 684
- gi No. 4262183
- Description: (AC005508) 51434 [Arabidopsis thaliana]
- % Identity: 95.9
- Alignment Length: 193
- Location of Alignment in SEQ ID NO 1460: from 34 to 220

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1461
- Ceres seq_id 1570071
- Location of start within SEQ ID NO 1459: at 350 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1461: from 1 to 94 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 685
- gi No. 4262183
- Description: (AC005508) 51434 [Arabidopsis thaliana]
- % Identity: 95.9
- Alignment Length: 193
- Location of Alignment in SEQ ID NO 1461: from 1 to 126

Maximum Length Sequence:
related to:
Clone IDs:
19110
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1462
- Ceres seq_id 1570072

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1463
- Ceres seq_id 1570073
- Location of start within SEQ ID NO 1462: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1463: from 84 to 132 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1464
- Ceres seq_id 1570074
- Location of start within SEQ ID NO 1462: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1464: from 59 to 107 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1465
- Ceres seq_id 1570075
- Location of start within SEQ ID NO 1462: at 298 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Kelch motif
- Location within SEQ ID NO 1465: from 7 to 55 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
19223
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1466
- Ceres seq_id 1570080
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1467
- Ceres seq_id 1570081
- Location of start within SEQ ID NO 1466: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L36e
- Location within SEQ ID NO 1467: from 18 to 74 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1468
- Ceres seq_id 1570082
- Location of start within SEQ ID NO 1466: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L36e
- Location within SEQ ID NO 1468: from 2 to 58 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
19740
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1469
- Ceres seq_id 1570107
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1470
- Ceres seq_id 1570108
- Location of start within SEQ ID NO 1469: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1471
- Ceres seq_id 1570109
- Location of start within SEQ ID NO 1469: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1472
- Ceres seq_id 1570110
- Location of start within SEQ ID NO 1469: at 401 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AN1-like Zinc finger
- Location within SEQ ID NO 1472: from 31 to 61 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 686
- gi No. 2148157
- Description: ubiquitin homolog - bovine (fragment) >gi|998680|bbs|165969 (S77733) ubiquitin homolog [cattle, endothelial cells, Peptide Partial, 46 aa] [Bos taurus]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 1472: from 45 to 55

Maximum Length Sequence:
related to:
Clone IDs:
1984

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1473
- Ceres seq_id 1570111

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1474
- Ceres seq_id 1570112
- Location of start within SEQ ID NO 1473: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 687
- gi No. 3914117
- Description: NUCLEOSIDE DIPHOSPHATE KINASE II PRECURSOR (NDK II) (NDP KINASE II) >gi|3093480 (AF017640) nucleoside diphosphate kinase type 2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 1474: from 6 to 74

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1475
- Ceres seq_id 1570113
- Location of start within SEQ ID NO 1473: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 688

- gi No. 3914117
- Description: NUCLEOSIDE DIPHOSPHATE KINASE II PRECURSOR (NDK II) (NDP KINASE II) >gi|3093480 (AF017640) nucleoside diphosphate kinase type 2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 1475: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
116619
19986
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1476
- Ceres seq_id 1570116
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1477
- Ceres seq_id 1570117
- Location of start within SEQ ID NO 1476: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 1477: from 39 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 689
- gi No. 1710591
- Description: 60S ACIDIC RIBOSOMAL PROTEIN P2
- % Identity: 96.2
- Alignment Length: 78
- Location of Alignment in SEQ ID NO 1477: from 38 to 115

Maximum Length Sequence:
related to:
Clone IDs:
20380
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1478
- Ceres seq_id 1570138
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1479
- Ceres seq_id 1570139
- Location of start within SEQ ID NO 1478: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 690
- gi No. 2388562
- Description: (AC000098) EST gb|ATTS0295 comes from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 1479: from 1 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1480
- Ceres seq_id 1570140
- Location of start within SEQ ID NO 1478: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 691
- gi No. 2388562
- Description: (AC000098) EST gb|ATTS0295 comes from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 1480: from 1 to 55

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1481
- Ceres seq_id 1570141
- Location of start within SEQ ID NO 1478: at 361 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20383
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1482
- Ceres seq_id 1570142
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1483
- Ceres seq_id 1570143
- Location of start within SEQ ID NO 1482: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1483: from 28 to 110 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20822
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1484
- Ceres seq_id 1570159
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1485
- Ceres seq_id 1570160
- Location of start within SEQ ID NO 1484: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 1485: from 58 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 692
- gi No. 2407802
- Description: (Y12576) histone H2B [Arabidopsis thaliana]

- % Identity: 96.7
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 1485: from 27 to 176

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1486
- Ceres seq_id 1570161
- Location of start within SEQ ID NO 1484: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 1486: from 32 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 693
- gi No. 2407802
- Description: (Y12576) histone H2B [Arabidopsis thaliana]
- % Identity: 96.7
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 1486: from 1 to 150

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1487
- Ceres seq_id 1570162
- Location of start within SEQ ID NO 1484: at 242 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 1487: from 1 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 694
- gi No. 2407802
- Description: (Y12576) histone H2B [Arabidopsis thaliana]
- % Identity: 96.7
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 1487: from 1 to 96

Maximum Length Sequence:
related to:
Clone IDs:
20837
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1488
- Ceres seq_id 1570163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1489
- Ceres seq_id 1570164
- Location of start within SEQ ID NO 1488: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1489: from 32 to 142 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1490

- Ceres seq_id 1570165
- Location of start within SEQ ID NO 1488: at 513 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1490: from 1 to 106 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1491
- Ceres seq_id 1570166
- Location of start within SEQ ID NO 1488: at 528 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 1491: from 1 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20946
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1492
- Ceres seq_id 1570170

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1493
- Ceres seq_id 1570171
- Location of start within SEQ ID NO 1492: at 248 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1493: from 171 to 239 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1494
- Ceres seq_id 1570172
- Location of start within SEQ ID NO 1492: at 545 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1494: from 72 to 140 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1495
- Ceres seq_id 1570173
- Location of start within SEQ ID NO 1492: at 707 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)

- Location within SEQ ID NO 1495: from 18 to 86 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
21223
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1496
- Ceres seq_id 1570183
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1497
- Ceres seq_id 1570184
- Location of start within SEQ ID NO 1496: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 695
- gi No. 4646206
- Description: (AC007230) Contains similarity to gb|D13630 KIAA0005 gene from Homo sapiens. ESTs gb|T45345, gb|T21086, gb|R90360, gb|T20468, gb|T45191 and gb|AI100459 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 287
- Location of Alignment in SEQ ID NO 1497: from 2 to 288

Maximum Length Sequence:
related to:
Clone IDs:
21618
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1498
- Ceres seq_id 1570189
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1499
- Ceres seq_id 1570190
- Location of start within SEQ ID NO 1498: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 696
- gi No. 1825645
- Description: (U88173) weak similarity to Arabidopsis thaliana ubiquitin-like protein 8 [Caenorhabditis elegans]
- % Identity: 78.9
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 1499: from 24 to 42

Maximum Length Sequence:
related to:
Clone IDs:
21868
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1500
- Ceres seq_id 1570201
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1501
- Ceres seq_id 1570202

- Location of start within SEQ ID NO 1500: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 1501: from 121 to 182 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 697
- gi No. 2281629
- Description: (AF003095) AP2 domain containing protein RAP2.2 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 1501: from 130 to 375

Maximum Length Sequence:
related to:
Clone IDs:
21999
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1502
- Ceres seq_id 1570227

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1503
- Ceres seq_id 1570228
- Location of start within SEQ ID NO 1502: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 1503: from 30 to 77 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1504
- Ceres seq_id 1570229
- Location of start within SEQ ID NO 1502: at 40 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 1504: from 17 to 64 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1505
- Ceres seq_id 1570230
- Location of start within SEQ ID NO 1502: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 1505: from 5 to 52 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
33482
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1506
- Ceres seq_id 1570235
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1507
- Ceres seq_id 1570236
- Location of start within SEQ ID NO 1506: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- MSP (Major sperm protein) domain.
- Location within SEQ ID NO 1507: from 7 to 145 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 698
- gi No. 3193330
- Description: (AF069299) contains similarity to Medicago sativa corC (GB:L22305) [Arabidopsis thaliana]
- % Identity: 86.1
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1507: from 1 to 250

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1508
- Ceres seq_id 1570237
- Location of start within SEQ ID NO 1506: at 182 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- MSP (Major sperm protein) domain.
- Location within SEQ ID NO 1508: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 699
- gi No. 3193330
- Description: (AF069299) contains similarity to Medicago sativa corC (GB:L22305) [Arabidopsis thaliana]
- % Identity: 86.1
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1508: from 1 to 223

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1509
- Ceres seq_id 1570238
- Location of start within SEQ ID NO 1506: at 314 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- MSP (Major sperm protein) domain.
- Location within SEQ ID NO 1509: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 700
- gi No. 3193330
- Description: (AF069299) contains similarity to Medicago sativa corC (GB:L22305) [Arabidopsis thaliana]

- % Identity: 86.1
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1509: from 1 to 179

Maximum Length Sequence:
related to:
Clone IDs:
39621
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1510
- Ceres seq_id 1570239
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1511
- Ceres seq_id 1570240
- Location of start within SEQ ID NO 1510: at 172 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 701
- gi No. 809345
- Description: Papain (E.C.4.3.22.2) Complex With E64-C (Form Ii)
- % Identity: 70
- Alignment Length: 30
- Location of Alignment in SEQ ID NO 1511: from 81 to 110

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1512
- Ceres seq_id 1570241
- Location of start within SEQ ID NO 1510: at 720 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 1512: from 1 to 116 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1513
- Ceres seq_id 1570242
- Location of start within SEQ ID NO 1510: at 780 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 1513: from 1 to 96 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
205622
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1514
- Ceres seq_id 1570247
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1515
- Ceres seq_id 1570248
- Location of start within SEQ ID NO 1514: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1515: from 116 to 292 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1516
- Ceres seq_id 1570249
- Location of start within SEQ ID NO 1514: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1516: from 105 to 281 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1517
- Ceres seq_id 1570250
- Location of start within SEQ ID NO 1514: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1517: from 86 to 262 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
206237
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1518
- Ceres seq_id 1570265
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1519
- Ceres seq_id 1570266
- Location of start within SEQ ID NO 1518: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 1519: from 70 to 123 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1520
- Ceres seq_id 1570267
- Location of start within SEQ ID NO 1518: at 872 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1521
- Ceres seq_id 1570268
- Location of start within SEQ ID NO 1518: at 995 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
207075
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1522
- Ceres seq_id 1570289

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1523
- Ceres seq_id 1570290
- Location of start within SEQ ID NO 1522: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 1523: from 130 to 196 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1524
- Ceres seq_id 1570291
- Location of start within SEQ ID NO 1522: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 1524: from 115 to 181 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1525
- Ceres seq_id 1570292
- Location of start within SEQ ID NO 1522: at 210 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 1525: from 61 to 127 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
207148
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1526
- Ceres seq_id 1570301

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1527
- Ceres seq_id 1570302

- Location of start within SEQ ID NO 1526: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1528
- Ceres seq_id 1570303
- Location of start within SEQ ID NO 1526: at 240 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1528: from 46 to 89 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
23674
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1529
- Ceres seq_id 1570308

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1530
- Ceres seq_id 1570309
- Location of start within SEQ ID NO 1529: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 1530: from 73 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 702
- gi No. 119975
- Description: FERREDOXIN PRECURSOR >gi|99692|pir||S09979 ferredoxin [2Fe-2S] precursor - Arabidopsis thaliana >gi|16437|emb|CAA35754| (X51370) ferredoxin precursor [Arabidopsis thaliana] >gi|166698 (M35868) ferrodoxin A [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 1530: from 18 to 165

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1531
- Ceres seq_id 1570310
- Location of start within SEQ ID NO 1529: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 1531: from 56 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 703
- gi No. 119975
- Description: FERREDOXIN PRECURSOR >gi|99692|pir||S09979 ferredoxin [2Fe-2S] precursor - Arabidopsis thaliana >gi|16437|emb|CAA35754|

(X51370) ferredoxin precursor [Arabidopsis thaliana] >gi|166698 (M35868) ferrodoxin A [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 1531: from 1 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1532
- Ceres seq_id 1570311
- Location of start within SEQ ID NO 1529: at 206 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 1532: from 5 to 89 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 704
- gi No. 119975
- Description: FERREDOXIN PRECURSOR >gi|99692|pir||S09979 ferredoxin [2Fe-2S] precursor - Arabidopsis thaliana >gi|16437|emb|CAA35754| (X51370) ferredoxin precursor [Arabidopsis thaliana] >gi|166698 (M35868) ferrodoxin A [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 1532: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
2975
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1533
- Ceres seq_id 1570315

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1534
- Ceres seq_id 1570316
- Location of start within SEQ ID NO 1533: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Osteopontin
- Location within SEQ ID NO 1534: from 42 to 186 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 705
- gi No. 1658271
- Description: (U74622) nascent polypeptide associated complex alpha chain [Nicotiana tabacum]
- % Identity: 94.1
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 1534: from 179 to 211

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1535
- Ceres seq_id 1570317
- Location of start within SEQ ID NO 1533: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Osteopontin

- Location within SEQ ID NO 1535: from 26 to 170 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 706
- gi No. 1658271
- Description: (U74622) nascent polypeptide associated complex alpha chain [Nicotiana tabacum]
- % Identity: 94.1
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 1535: from 163 to 195

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1536
- Ceres seq_id 1570318
- Location of start within SEQ ID NO 1533: at 257 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Osteopontin
- Location within SEQ ID NO 1536: from 1 to 101 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 707
- gi No. 1658271
- Description: (U74622) nascent polypeptide associated complex alpha chain [Nicotiana tabacum]
- % Identity: 94.1
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 1536: from 94 to 126

Maximum Length Sequence:
related to:
Clone IDs:
32449
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1537
- Ceres seq_id 1570339
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1538
- Ceres seq_id 1570340
- Location of start within SEQ ID NO 1537: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 1538: from 9 to 341 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 708
- gi No. 1170373
- Description: HEAT SHOCK COGNATE 70 KD PROTEIN 1 >gi|1072473|pir||S46302 heat shock cognate protein 70-1 - Arabidopsis thaliana >gi|397482|emb|CAA52684| (X74604) heat shock protein 70 cognate [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 1538: from 1 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1539
- Ceres seq_id 1570341

- Location of start within SEQ ID NO 1537: at 295 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 1539: from 1 to 278 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 709
- gi No. 1170373
- Description: HEAT SHOCK COGNATE 70 KD PROTEIN 1 >gi|1072473|pir||S46302 heat shock cognate protein 70-1 - Arabidopsis thaliana >gi|397482|emb|CAA52684| (X74604) heat shock protein 70 cognate [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 1539: from 1 to 278

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1540
- Ceres seq_id 1570342
- Location of start within SEQ ID NO 1537: at 373 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 1540: from 1 to 252 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 710
- gi No. 1170373
- Description: HEAT SHOCK COGNATE 70 KD PROTEIN 1 >gi|1072473|pir||S46302 heat shock cognate protein 70-1 - Arabidopsis thaliana >gi|397482|emb|CAA52684| (X74604) heat shock protein 70 cognate [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 1540: from 1 to 252

Maximum Length Sequence:
related to:
Clone IDs:
8188

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1541
- Ceres seq_id 1570343

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1542
- Ceres seq_id 1570344
- Location of start within SEQ ID NO 1541: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1542: from 82 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 711
- gi No. 2129603
- Description: glycine/proline-rich protein - Arabidopsis thaliana >gi|1465364|emb|CAA59059| (X84315) GPRP [Arabidopsis thaliana]

- % Identity: 72.1
- Alignment Length: 199
- Location of Alignment in SEQ ID NO 1542: from 1 to 179

Maximum Length Sequence:
related to:
Clone IDs:
31105
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1543
- Ceres seq_id 1570349
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1544
- Ceres seq_id 1570350
- Location of start within SEQ ID NO 1543: at 1234 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1544: from 78 to 272 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1545
- Ceres seq_id 1570351
- Location of start within SEQ ID NO 1543: at 1477 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1545: from 1 to 191 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
8462
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1546
- Ceres seq_id 1570356
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1547
- Ceres seq_id 1570357
- Location of start within SEQ ID NO 1546: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 1547: from 119 to 186 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1548
- Ceres seq_id 1570358
- Location of start within SEQ ID NO 1546: at 245 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 1548: from 59 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1549
- Ceres seq_id 1570359
- Location of start within SEQ ID NO 1546: at 341 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 1549: from 27 to 94 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32443
125396
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1550
- Ceres seq_id 1570366
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1551
- Ceres seq_id 1570367
- Location of start within SEQ ID NO 1550: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- PHD-finger
- Location within SEQ ID NO 1551: from 243 to 295 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 712
- gi No. 3319341
- Description: (AF077407) similar to Medicago sativa nucleic acid binding protein Alfin-1 (GB:L07291) [Arabidopsis thaliana]
- % Identity: 94.4
- Alignment Length: 259
- Location of Alignment in SEQ ID NO 1551: from 45 to 296

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1552
- Ceres seq_id 1570368
- Location of start within SEQ ID NO 1550: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- PHD-finger
- Location within SEQ ID NO 1552: from 199 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 713
- gi No. 3319341
- Description: (AF077407) similar to Medicago sativa nucleic acid binding protein Alfin-1 (GB:L07291) [Arabidopsis thaliana]
- % Identity: 94.4
- Alignment Length: 259

- Location of Alignment in SEQ ID NO 1552: from 1 to 252

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1553
- Ceres seq_id 1570369
- Location of start within SEQ ID NO 1550: at 209 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- PHD-finger
- Location within SEQ ID NO 1553: from 174 to 226 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 714
- gi No. 3319341
- Description: (AF077407) similar to Medicago sativa nucleic acid binding protein Alfin-1 (GB:L07291) [Arabidopsis thaliana]
- % Identity: 94.4
- Alignment Length: 259
- Location of Alignment in SEQ ID NO 1553: from 1 to 227

Maximum Length Sequence:
related to:
Clone IDs:
29659
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1554
- Ceres seq_id 1570370
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1555
- Ceres seq_id 1570371
- Location of start within SEQ ID NO 1554: at 187 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1555: from 161 to 201 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32753
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1556
- Ceres seq_id 1570372
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1557
- Ceres seq_id 1570373
- Location of start within SEQ ID NO 1556: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L18p family
- Location within SEQ ID NO 1557: from 13 to 229 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 715
- gi No. 3915826
- Description: 60S RIBOSOMAL PROTEIN L5

- % Identity: 77.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1557: from 3 to 284

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1558
- Ceres seq_id 1570374
- Location of start within SEQ ID NO 1556: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L18p family
- Location within SEQ ID NO 1558: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 716
- gi No. 3915826
- Description: 60S RIBOSOMAL PROTEIN L5
- % Identity: 77.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1558: from 1 to 173

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1559
- Ceres seq_id 1570375
- Location of start within SEQ ID NO 1556: at 405 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L18p family
- Location within SEQ ID NO 1559: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 717
- gi No. 3915826
- Description: 60S RIBOSOMAL PROTEIN L5
- % Identity: 77.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 1559: from 1 to 170

Maximum Length Sequence:
related to:
Clone IDs:
31579
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1560
- Ceres seq_id 1570376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1561
- Ceres seq_id 1570377
- Location of start within SEQ ID NO 1560: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1561: from 1 to 319 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1562

- Ceres seq_id 1570378
- Location of start within SEQ ID NO 1560: at 368 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1562: from 1 to 318 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1563
- Ceres seq_id 1570379
- Location of start within SEQ ID NO 1560: at 401 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 1563: from 1 to 307 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34406
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1564
- Ceres seq_id 1570380
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1565
- Ceres seq_id 1570381
- Location of start within SEQ ID NO 1564: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1565: from 39 to 110 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1566
- Ceres seq_id 1570382
- Location of start within SEQ ID NO 1564: at 408 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1567
- Ceres seq_id 1570383
- Location of start within SEQ ID NO 1564: at 444 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
37292
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1568
- Ceres seq_id 1570384
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1569
- Ceres seq_id 1570385
- Location of start within SEQ ID NO 1568: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 1569: from 27 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 718
- gi No. 1362106
- Description: GUT 7-2a protein - common tobacco (fragment)
- % Identity: 81.1
- Alignment Length: 90
- Location of Alignment in SEQ ID NO 1569: from 33 to 122

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1570
- Ceres seq_id 1570386
- Location of start within SEQ ID NO 1568: at 495 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 719
- gi No. 1362106
- Description: GUT 7-2a protein - common tobacco (fragment)
- % Identity: 81.1
- Alignment Length: 90
- Location of Alignment in SEQ ID NO 1570: from 1 to 12

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1571
- Ceres seq_id 1570387
- Location of start within SEQ ID NO 1568: at 501 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
29661
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1572
- Ceres seq_id 1570388
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1573
- Ceres seq_id 1570389
- Location of start within SEQ ID NO 1572: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Cytochrome P450
- Location within SEQ ID NO 1573: from 83 to 512 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1574
- Ceres seq_id 1570390
- Location of start within SEQ ID NO 1572: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1574: from 67 to 496 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1575
- Ceres seq_id 1570391
- Location of start within SEQ ID NO 1572: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1575: from 1 to 418 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32857
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1576
- Ceres seq_id 1570412
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1577
- Ceres seq_id 1570413
- Location of start within SEQ ID NO 1576: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 720
- gi No. 2623304
- Description: (AC002409) similar to Medicago nodulin N21 [Arabidopsis thaliana]
- % Identity: 80.6
- Alignment Length: 160
- Location of Alignment in SEQ ID NO 1577: from 1 to 160

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1578
- Ceres seq_id 1570414
- Location of start within SEQ ID NO 1576: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 721

- gi No. 2623304
- Description: (AC002409) similar to Medicago nodulin N21 [Arabidopsis thaliana]
- % Identity: 80.6
- Alignment Length: 160
- Location of Alignment in SEQ ID NO 1578: from 1 to 156

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1579
- Ceres seq_id 1570415
- Location of start within SEQ ID NO 1576: at 759 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Integral membrane protein
- Location within SEQ ID NO 1579: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
7468
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1580
- Ceres seq_id 1570422

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1581
- Ceres seq_id 1570423
- Location of start within SEQ ID NO 1580: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1581: from 107 to 176 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1582
- Ceres seq_id 1570424
- Location of start within SEQ ID NO 1580: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1582: from 91 to 160 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1583
- Ceres seq_id 1570425
- Location of start within SEQ ID NO 1580: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1583: from 64 to 133 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
23082
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1584
- Ceres seq_id 1570437
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1585
- Ceres seq_id 1570438
- Location of start within SEQ ID NO 1584: at 123 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 1585: from 96 to 156 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1586
- Ceres seq_id 1570439
- Location of start within SEQ ID NO 1584: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 1586: from 86 to 146 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1587
- Ceres seq_id 1570440
- Location of start within SEQ ID NO 1584: at 336 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 1587: from 25 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24140
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1588
- Ceres seq_id 1570446
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1589
- Ceres seq_id 1570447
- Location of start within SEQ ID NO 1588: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rhodanese-like domain
- Location within SEQ ID NO 1589: from 18 to 121 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1590
- Ceres seq_id 1570448
- Location of start within SEQ ID NO 1588: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rhodanese-like domain
- Location within SEQ ID NO 1590: from 1 to 65 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
23231
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1591
- Ceres seq_id 1570449

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1592
- Ceres seq_id 1570450
- Location of start within SEQ ID NO 1591: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 722
- gi No. 2213616
- Description: (AC000103) F21J9.10 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 1592: from 7 to 76

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1593
- Ceres seq_id 1570451
- Location of start within SEQ ID NO 1591: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 723
- gi No. 2213616
- Description: (AC000103) F21J9.10 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 1593: from 1 to 70

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1594
- Ceres seq_id 1570452
- Location of start within SEQ ID NO 1591: at 352 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1594: from 75 to 111 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 724
- gi No. 2213616
- Description: (AC000103) F21J9.10 [Arabidopsis thaliana]
- % Identity: 80.3
- Alignment Length: 214
- Location of Alignment in SEQ ID NO 1594: from 1 to 116

Maximum Length Sequence:
related to:
Clone IDs:
34999
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1595
- Ceres seq_id 1570474
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1596
- Ceres seq_id 1570475
- Location of start within SEQ ID NO 1595: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 725
- gi No. 1946362
- Description: (U93215) photosystem II reaction center 6.1KD protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 133
- Location of Alignment in SEQ ID NO 1596: from 1 to 133

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1597
- Ceres seq_id 1570476
- Location of start within SEQ ID NO 1595: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 726
- gi No. 1946362
- Description: (U93215) photosystem II reaction center 6.1KD protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 133
- Location of Alignment in SEQ ID NO 1597: from 1 to 98

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1598
- Ceres seq_id 1570477
- Location of start within SEQ ID NO 1595: at 280 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 727
- gi No. 1946362
- Description: (U93215) photosystem II reaction center 6.1KD protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 133

- Location of Alignment in SEQ ID NO 1598: from 1 to 87

Maximum Length Sequence:
related to:
Clone IDs:
26126
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1599
- Ceres seq_id 1570478
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1600
- Ceres seq_id 1570479
- Location of start within SEQ ID NO 1599: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1600: from 2 to 217 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1601
- Ceres seq_id 1570480
- Location of start within SEQ ID NO 1599: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1601: from 1 to 177 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1602
- Ceres seq_id 1570481
- Location of start within SEQ ID NO 1599: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1602: from 1 to 176 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
27304
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1603
- Ceres seq_id 1570485
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1604
- Ceres seq_id 1570486
- Location of start within SEQ ID NO 1603: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 1604: from 46 to 146 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1605
- Ceres seq_id 1570487
- Location of start within SEQ ID NO 1603: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 1605: from 44 to 144 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1606
- Ceres seq_id 1570488
- Location of start within SEQ ID NO 1603: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 1606: from 39 to 139 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24775
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1607
- Ceres seq_id 1570489

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1608
- Ceres seq_id 1570490
- Location of start within SEQ ID NO 1607: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isochorismatase family
- Location within SEQ ID NO 1608: from 26 to 182 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1609
- Ceres seq_id 1570491
- Location of start within SEQ ID NO 1607: at 54 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isochorismatase family
- Location within SEQ ID NO 1609: from 9 to 165 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1610
- Ceres seq_id 1570492

- Location of start within SEQ ID NO 1607: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Isochorismatase family
- Location within SEQ ID NO 1610: from 1 to 151 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24286
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1611
- Ceres seq_id 1570493
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1612
- Ceres seq_id 1570494
- Location of start within SEQ ID NO 1611: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1612: from 30 to 104 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1613
- Ceres seq_id 1570495
- Location of start within SEQ ID NO 1611: at 11 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1613: from 27 to 101 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1614
- Ceres seq_id 1570496
- Location of start within SEQ ID NO 1611: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1614: from 20 to 94 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24849
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1615
- Ceres seq_id 1570497
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1616

- Ceres seq_id 1570498
- Location of start within SEQ ID NO 1615: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- BTB/POZ domain
- Location within SEQ ID NO 1616: from 1 to 95 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1617
- Ceres seq_id 1570499
- Location of start within SEQ ID NO 1615: at 33 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- BTB/POZ domain
- Location within SEQ ID NO 1617: from 1 to 85 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1618
- Ceres seq_id 1570500
- Location of start within SEQ ID NO 1615: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- BTB/POZ domain
- Location within SEQ ID NO 1618: from 1 to 56 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
9683
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1619
- Ceres seq_id 1570501

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1620
- Ceres seq_id 1570502
- Location of start within SEQ ID NO 1619: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1620: from 25 to 208 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1621
- Ceres seq_id 1570503
- Location of start within SEQ ID NO 1619: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase

- Location within SEQ ID NO 1621: from 10 to 193 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1622
- Ceres seq_id 1570504
- Location of start within SEQ ID NO 1619: at 383 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 1622: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
27978
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1623
- Ceres seq_id 1570505
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1624
- Ceres seq_id 1570506
- Location of start within SEQ ID NO 1623: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L19e
- Location within SEQ ID NO 1624: from 23 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 728
- gi No. 548751
- Description: 60S RIBOSOMAL PROTEIN L19 >gi|1079135|pir||A61627 ribosomal protein L19 - fruit fly (Drosophila melanogaster) >gi|400448|emb|CAA52784| (X74776) ribosomal protein L19 [Drosophila melanogaster]
- % Identity: 71.9
- Alignment Length: 64
- Location of Alignment in SEQ ID NO 1624: from 20 to 83

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1625
- Ceres seq_id 1570507
- Location of start within SEQ ID NO 1623: at 341 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L19e
- Location within SEQ ID NO 1625: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 729
- gi No. 4506609
- Description: ref|NP_000972.1|pRPL19| ribosomal protein L19 >gi|132739|sp|P14118|RL19_HUMAN 60S RIBOSOMAL PROTEIN L19 >gi|71265|pir||R5RT19 ribosomal protein L19 - rat >gi|477424|pir||A48992 ribosomal protein L19 - human [Rattus norvegicus]

- % Identity: 70.8
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 1625: from 1 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1626
- Ceres seq_id 1570508
- Location of start within SEQ ID NO 1623: at 350 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L19e
- Location within SEQ ID NO 1626: from 1 to 51 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 730
- gi No. 4506609
- Description: ref|NP_000972.1|pRPL19| ribosomal protein L19 >gi|132739|sp|P14118|RL19_HUMAN 60S RIBOSOMAL PROTEIN L19 >gi|71265|pir||R5RT19 ribosomal protein L19 - rat >gi|477424|pir||A48992 ribosomal protein L19 - human [Rattus norvegicus]
- % Identity: 70.8
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 1626: from 1 to 83

Maximum Length Sequence:
related to:
Clone IDs:
28484
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1627
- Ceres seq_id 1570515

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1628
- Ceres seq_id 1570516
- Location of start within SEQ ID NO 1627: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 1628: from 23 to 133 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1629
- Ceres seq_id 1570517
- Location of start within SEQ ID NO 1627: at 213 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 1629: from 16 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1630
- Ceres seq_id 1570518
- Location of start within SEQ ID NO 1627: at 216 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 1630: from 15 to 125 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
23237
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1631
- Ceres seq_id 1570534
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1632
- Ceres seq_id 1570535
- Location of start within SEQ ID NO 1631: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ferritins
- Location within SEQ ID NO 1632: from 120 to 276 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 731
- gi No. 2970654
- Description: (AF052058) ferritin subunit cowpea2 precursor [Vigna unguiculata]
- % Identity: 71
- Alignment Length: 261
- Location of Alignment in SEQ ID NO 1632: from 23 to 281

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1633
- Ceres seq_id 1570536
- Location of start within SEQ ID NO 1631: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ferritins
- Location within SEQ ID NO 1633: from 98 to 254 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 732
- gi No. 2970654
- Description: (AF052058) ferritin subunit cowpea2 precursor [Vigna unguiculata]
- % Identity: 71
- Alignment Length: 261
- Location of Alignment in SEQ ID NO 1633: from 1 to 259

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1634
- Ceres seq_id 1570537
- Location of start within SEQ ID NO 1631: at 418 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ferritins
- Location within SEQ ID NO 1634: from 1 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 733
- gi No. 2970654
- Description: (AF052058) ferritin subunit cowpea2 precursor [Vigna unguiculata]
- % Identity: 71
- Alignment Length: 261
- Location of Alignment in SEQ ID NO 1634: from 1 to 142

Maximum Length Sequence:
related to:
Clone IDs:
2489
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1635
- Ceres seq_id 1570538

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1636
- Ceres seq_id 1570539
- Location of start within SEQ ID NO 1635: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1636: from 77 to 156 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1637
- Ceres seq_id 1570540
- Location of start within SEQ ID NO 1635: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1637: from 61 to 140 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1638
- Ceres seq_id 1570541
- Location of start within SEQ ID NO 1635: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1638: from 42 to 121 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3982
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1639
- Ceres seq_id 1570561

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1640
- Ceres seq_id 1570562
- Location of start within SEQ ID NO 1639: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1640: from 45 to 274 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 734
- gi No. 3451065
- Description: (AL031326) water channel-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 287
- Location of Alignment in SEQ ID NO 1640: from 1 to 287

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1641
- Ceres seq_id 1570563
- Location of start within SEQ ID NO 1639: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1641: from 1 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 735
- gi No. 3451065
- Description: (AL031326) water channel-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 287
- Location of Alignment in SEQ ID NO 1641: from 1 to 214

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1642
- Ceres seq_id 1570564
- Location of start within SEQ ID NO 1639: at 291 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1642: from 1 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 736
- gi No. 3451065
- Description: (AL031326) water channel-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 287
- Location of Alignment in SEQ ID NO 1642: from 1 to 206

Maximum Length Sequence:
related to:
Clone IDs:
6176
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 1643
- Ceres seq_id 1570565

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1644
- Ceres seq_id 1570566
- Location of start within SEQ ID NO 1643: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1644: from 21 to 214 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 737
- gi No. 2842494
- Description: (AL021749) prohibitin-like protein [Arabidopsis thaliana] >gi|4097688 (U66591) prohibitin 1 [Arabidopsis thaliana] >gi|4097694 (U66594) prohibitin 1 [Arabidopsis thaliana]
- % Identity: 88.1
- Alignment Length: 285
- Location of Alignment in SEQ ID NO 1644: from 5 to 285

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1645
- Ceres seq_id 1570567
- Location of start within SEQ ID NO 1643: at 281 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1645: from 1 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 738
- gi No. 2842494
- Description: (AL021749) prohibitin-like protein [Arabidopsis thaliana] >gi|4097688 (U66591) prohibitin 1 [Arabidopsis thaliana] >gi|4097694 (U66594) prohibitin 1 [Arabidopsis thaliana]
- % Identity: 88.1
- Alignment Length: 285
- Location of Alignment in SEQ ID NO 1645: from 1 to 222

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1646
- Ceres seq_id 1570568
- Location of start within SEQ ID NO 1643: at 371 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Band 7 family
- Location within SEQ ID NO 1646: from 1 to 121 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 739
- gi No. 2842494
- Description: (AL021749) prohibitin-like protein [Arabidopsis thaliana] >gi|4097688 (U66591) prohibitin 1 [Arabidopsis thaliana] >gi|4097694 (U66594) prohibitin 1 [Arabidopsis thaliana]
- % Identity: 88.1
- Alignment Length: 285
- Location of Alignment in SEQ ID NO 1646: from 1 to 192

Maximum Length Sequence:
related to:
Clone IDs:
29997
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1647
- Ceres seq_id 1570569
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1648
- Ceres seq_id 1570570
- Location of start within SEQ ID NO 1647: at 52 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S8e
- Location within SEQ ID NO 1648: from 3 to 127 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 740
- gi No. 1498053
- Description: (U64436) ribosomal protein S8 [Zea mays]
- % Identity: 83.7
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 1648: from 1 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1649
- Ceres seq_id 1570571
- Location of start within SEQ ID NO 1647: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S8e
- Location within SEQ ID NO 1649: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 741
- gi No. 1498053
- Description: (U64436) ribosomal protein S8 [Zea mays]
- % Identity: 83.7
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 1649: from 1 to 193

Maximum Length Sequence:
related to:
Clone IDs:
7632
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1650
- Ceres seq_id 1570576
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1651
- Ceres seq_id 1570577
- Location of start within SEQ ID NO 1650: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 742
- gi No. 2245000

- Description: (Z97341) LET1 like protein [Arabidopsis thaliana]
- % Identity: 98.2
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 1651: from 1 to 109

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1652
- Ceres seq_id 1570578
- Location of start within SEQ ID NO 1650: at 271 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 743
- gi No. 2245000
- Description: (Z97341) LET1 like protein [Arabidopsis thaliana]
- % Identity: 98.2
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 1652: from 1 to 66

Maximum Length Sequence:
related to:
Clone IDs:
27375
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1653
- Ceres seq_id 1570583

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1654
- Ceres seq_id 1570584
- Location of start within SEQ ID NO 1653: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 1654: from 31 to 188 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 744
- gi No. 2801444
- Description: (AF028339) ubiquitin-conjugating enzyme 16 [Arabidopsis thaliana]
- % Identity: 93.2
- Alignment Length: 162
- Location of Alignment in SEQ ID NO 1654: from 31 to 191

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1655
- Ceres seq_id 1570585
- Location of start within SEQ ID NO 1653: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 1655: from 1 to 158 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 745
- gi No. 2801444
- Description: (AF028339) ubiquitin-conjugating enzyme 16 [Arabidopsis thaliana]

- % Identity: 93.2
- Alignment Length: 162
- Location of Alignment in SEQ ID NO 1655: from 1 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1656
- Ceres seq_id 1570586
- Location of start within SEQ ID NO 1653: at 177 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 1656: from 1 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 746
- gi No. 2801444
- Description: (AF028339) ubiquitin-conjugating enzyme 16 [Arabidopsis thaliana]
- % Identity: 93.2
- Alignment Length: 162
- Location of Alignment in SEQ ID NO 1656: from 1 to 133

Maximum Length Sequence:
related to:
Clone IDs:
34408
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1657
- Ceres seq_id 1570603

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1658
- Ceres seq_id 1570604
- Location of start within SEQ ID NO 1657: at 14 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S6e
- Location within SEQ ID NO 1658: from 1 to 129 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 747
- gi No. 3281858
- Description: (AL031004) ribosomal protein S6 - like [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 250
- Location of Alignment in SEQ ID NO 1658: from 1 to 250

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1659
- Ceres seq_id 1570605
- Location of start within SEQ ID NO 1657: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S6e
- Location within SEQ ID NO 1659: from 1 to 67 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 748

- gi No. 3281858
- Description: (AL031004) ribosomal protein S6 - like [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 250
- Location of Alignment in SEQ ID NO 1659: from 1 to 188

Maximum Length Sequence:
related to:
Clone IDs:
33868
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1660
- Ceres seq_id 1570606
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1661
- Ceres seq_id 1570607
- Location of start within SEQ ID NO 1660: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S6e
- Location within SEQ ID NO 1661: from 10 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 749
- gi No. 3281858
- Description: (AL031004) ribosomal protein S6 - like [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 207
- Location of Alignment in SEQ ID NO 1661: from 10 to 216

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1662
- Ceres seq_id 1570608
- Location of start within SEQ ID NO 1660: at 12 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S6e
- Location within SEQ ID NO 1662: from 7 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 750
- gi No. 3281858
- Description: (AL031004) ribosomal protein S6 - like [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 207
- Location of Alignment in SEQ ID NO 1662: from 7 to 213

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1663
- Ceres seq_id 1570609
- Location of start within SEQ ID NO 1660: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S6e
- Location within SEQ ID NO 1663: from 1 to 67 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 751
- gi No. 3281858
- Description: (AL031004) ribosomal protein S6 - like [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 207
- Location of Alignment in SEQ ID NO 1663: from 1 to 188

Maximum Length Sequence:
related to:
Clone IDs:
26540
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1664
- Ceres seq_id 1570610
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1665
- Ceres seq_id 1570611
- Location of start within SEQ ID NO 1664: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 1665: from 63 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 752
- gi No. 2464879
- Description: (Z99707) caltractin-like protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 1665: from 1 to 158

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1666
- Ceres seq_id 1570612
- Location of start within SEQ ID NO 1664: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 1666: from 12 to 40 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 753
- gi No. 2464879
- Description: (Z99707) caltractin-like protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 1666: from 1 to 107

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1667
- Ceres seq_id 1570613
- Location of start within SEQ ID NO 1664: at 248 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 1667: from 5 to 33 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 754
- gi No. 2464879
- Description: (Z99707) caltractin-like protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 1667: from 1 to 100

Maximum Length Sequence:
related to:
Clone IDs:
25271
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1668
- Ceres seq_id 1570618
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1669
- Ceres seq_id 1570619
- Location of start within SEQ ID NO 1668: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 1669: from 7 to 93 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1670
- Ceres seq_id 1570620
- Location of start within SEQ ID NO 1668: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- CRAL/TRIO domain.
- Location within SEQ ID NO 1670: from 1 to 64 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1671
- Ceres seq_id 1570621
- Location of start within SEQ ID NO 1668: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
30978
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1672
- Ceres seq_id 1570622

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1673
- Ceres seq_id 1570623
- Location of start within SEQ ID NO 1672: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L30
- Location within SEQ ID NO 1673: from 105 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 755
- gi No. 445613
- Description: ribosomal protein L7 [Solanum tuberosum]
- % Identity: 79.3
- Alignment Length: 111
- Location of Alignment in SEQ ID NO 1673: from 24 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1674
- Ceres seq_id 1570624
- Location of start within SEQ ID NO 1672: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L30
- Location within SEQ ID NO 1674: from 82 to 124 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 756
- gi No. 445613
- Description: ribosomal protein L7 [Solanum tuberosum]
- % Identity: 79.3
- Alignment Length: 111
- Location of Alignment in SEQ ID NO 1674: from 1 to 111

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1675
- Ceres seq_id 1570625
- Location of start within SEQ ID NO 1672: at 621 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 757
- gi No. 445613
- Description: ribosomal protein L7 [Solanum tuberosum]
- % Identity: 90.2
- Alignment Length: 133
- Location of Alignment in SEQ ID NO 1675: from 1 to 117

Maximum Length Sequence:
related to:
Clone IDs:
144620
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1676
- Ceres seq_id 1570636

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1677
- Ceres seq_id 1570637

- Location of start within SEQ ID NO 1676: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 1677: from 144 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 758
- gi No. 548653
- Description: 50S RIBOSOMAL PROTEIN L12-A, CHLOROPLAST PRECURSOR (CL12-A) >gi|541895|pir||A53394 ribosomal protein L12.A, chloroplast - Arabidopsis thaliana >gi|468771|emb|CAA48181| (X68046) ribosomal protein L12 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 191
- Location of Alignment in SEQ ID NO 1677: from 22 to 212

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1678
- Ceres seq_id 1570638
- Location of start within SEQ ID NO 1676: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 1678: from 123 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 759
- gi No. 548653
- Description: 50S RIBOSOMAL PROTEIN L12-A, CHLOROPLAST PRECURSOR (CL12-A) >gi|541895|pir||A53394 ribosomal protein L12.A, chloroplast - Arabidopsis thaliana >gi|468771|emb|CAA48181| (X68046) ribosomal protein L12 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 191
- Location of Alignment in SEQ ID NO 1678: from 1 to 191

Maximum Length Sequence:
related to:
Clone IDs:
40474
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1679
- Ceres seq_id 1570655

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1680
- Ceres seq_id 1570656
- Location of start within SEQ ID NO 1679: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1680: from 44 to 273 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 760
- gi No. 1175012

- Description: PLASMA MEMBRANE INTRINSIC PROTEIN 1C (TRANSMEMBRANE PROTEIN B) (TMP-B) >gi|396218|emb|CAA49155| (X69294) transmembrane protein TMP-B [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 1680: from 1 to 286

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1681
- Ceres seq_id 1570657
- Location of start within SEQ ID NO 1679: at 279 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1681: from 1 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 761
- gi No. 1175012
- Description: PLASMA MEMBRANE INTRINSIC PROTEIN 1C (TRANSMEMBRANE PROTEIN B) (TMP-B) >gi|396218|emb|CAA49155| (X69294) transmembrane protein TMP-B [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 1681: from 1 to 214

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1682
- Ceres seq_id 1570658
- Location of start within SEQ ID NO 1679: at 303 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 1682: from 1 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 762
- gi No. 1175012
- Description: PLASMA MEMBRANE INTRINSIC PROTEIN 1C (TRANSMEMBRANE PROTEIN B) (TMP-B) >gi|396218|emb|CAA49155| (X69294) transmembrane protein TMP-B [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 286
- Location of Alignment in SEQ ID NO 1682: from 1 to 206

Maximum Length Sequence:
related to:
Clone IDs:
41439
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1683
- Ceres seq_id 1570669
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1684
- Ceres seq_id 1570670
- Location of start within SEQ ID NO 1683: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Zinc finger, C2H2 type
- Location within SEQ ID NO 1684: from 165 to 187 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 763
- gi No. 1786138
- Description: (AB000453) PEThy;ZPT3-1 [Petunia x hybrida]
- % Identity: 73.1
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 1684: from 163 to 188

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1685
- Ceres seq_id 1570671
- Location of start within SEQ ID NO 1683: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 1685: from 160 to 182 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 764
- gi No. 1786138
- Description: (AB000453) PEThy;ZPT3-1 [Petunia x hybrida]
- % Identity: 73.1
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 1685: from 158 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1686
- Ceres seq_id 1570672
- Location of start within SEQ ID NO 1683: at 190 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 1686: from 143 to 165 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 765
- gi No. 1786138
- Description: (AB000453) PEThy;ZPT3-1 [Petunia x hybrida]
- % Identity: 73.1
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 1686: from 141 to 166

Maximum Length Sequence:
related to:
Clone IDs:
38063
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1687
- Ceres seq_id 1570681

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1688
- Ceres seq_id 1570682
- Location of start within SEQ ID NO 1687: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Ribosomal protein S5
- Location within SEQ ID NO 1688: from 177 to 306 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 766
- gi No. 1707008
- Description: (U78721) 30S ribosomal protein S5 isolog [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1688: from 108 to 329

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1689
- Ceres seq_id 1570683
- Location of start within SEQ ID NO 1687: at 424 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S5
- Location within SEQ ID NO 1689: from 36 to 165 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 767
- gi No. 1707008
- Description: (U78721) 30S ribosomal protein S5 isolog [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1689: from 1 to 188

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1690
- Ceres seq_id 1570684
- Location of start within SEQ ID NO 1687: at 448 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S5
- Location within SEQ ID NO 1690: from 28 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 768
- gi No. 1707008
- Description: (U78721) 30S ribosomal protein S5 isolog [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 222
- Location of Alignment in SEQ ID NO 1690: from 1 to 180

Maximum Length Sequence:
related to:
Clone IDs:
38383
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1691
- Ceres seq_id 1570692

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1692
- Ceres seq_id 1570693
- Location of start within SEQ ID NO 1691: at 141 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1692: from 260 to 307 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 769
- gi No. 2864602
- Description: (Y12071) thylakoid lumen rotamase [Spinacia oleracea]
- % Identity: 76.4
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1692: from 33 to 437

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1693
- Ceres seq_id 1570694
- Location of start within SEQ ID NO 1691: at 342 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1693: from 193 to 240 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 770
- gi No. 2864602
- Description: (Y12071) thylakoid lumen rotamase [Spinacia oleracea]
- % Identity: 76.4
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1693: from 1 to 370

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1694
- Ceres seq_id 1570695
- Location of start within SEQ ID NO 1691: at 666 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 1694: from 85 to 132 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 771
- gi No. 2864602
- Description: (Y12071) thylakoid lumen rotamase [Spinacia oleracea]
- % Identity: 76.4
- Alignment Length: 406
- Location of Alignment in SEQ ID NO 1694: from 1 to 262

Maximum Length Sequence:
related to:
Clone IDs:
3542
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1695
- Ceres seq_id 1570696
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1696
- Ceres seq_id 1570697
- Location of start within SEQ ID NO 1695: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger C-x8-C-x5-C-x3-H type (and similar).
- Location within SEQ ID NO 1696: from 271 to 300 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1697
- Ceres seq_id 1570698
- Location of start within SEQ ID NO 1695: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger C-x8-C-x5-C-x3-H type (and similar).
- Location within SEQ ID NO 1697: from 235 to 264 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1698
- Ceres seq_id 1570699
- Location of start within SEQ ID NO 1695: at 162 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger C-x8-C-x5-C-x3-H type (and similar).
- Location within SEQ ID NO 1698: from 218 to 247 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24619
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1699
- Ceres seq_id 1570716

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1700
- Ceres seq_id 1570717
- Location of start within SEQ ID NO 1699: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1700: from 19 to 85 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1701
- Ceres seq_id 1570718
- Location of start within SEQ ID NO 1699: at 141 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Sm protein
- Location within SEQ ID NO 1701: from 10 to 76 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
24640
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1702
- Ceres seq_id 1570719
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1703
- Ceres seq_id 1570720
- Location of start within SEQ ID NO 1702: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 772
- gi No. 1064887
- Description: (X92955) pollen coat protein [Brassica oleracea]
- % Identity: 93.8
- Alignment Length: 64
- Location of Alignment in SEQ ID NO 1703: from 30 to 93

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1704
- Ceres seq_id 1570721
- Location of start within SEQ ID NO 1702: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 773
- gi No. 1064887
- Description: (X92955) pollen coat protein [Brassica oleracea]
- % Identity: 93.8
- Alignment Length: 64
- Location of Alignment in SEQ ID NO 1704: from 1 to 60

Maximum Length Sequence:
related to:
Clone IDs:
25204
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1705
- Ceres seq_id 1570735
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1706
- Ceres seq_id 1570736
- Location of start within SEQ ID NO 1705: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 774
- gi No. 3769300
- Description: (AB018552) AtRER1C [Arabidopsis thaliana]
- % Identity: 100

- Alignment Length: 212
- Location of Alignment in SEQ ID NO 1706: from 1 to 212

Maximum Length Sequence:
related to:
Clone IDs:
25429
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1707
- Ceres seq_id 1570737
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1708
- Ceres seq_id 1570738
- Location of start within SEQ ID NO 1707: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 1708: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 775
- gi No. 1710591
- Description: 60S ACIDIC RIBOSOMAL PROTEIN P2
- % Identity: 76.6
- Alignment Length: 77
- Location of Alignment in SEQ ID NO 1708: from 39 to 113

Maximum Length Sequence:
related to:
Clone IDs:
2617
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1709
- Ceres seq_id 1570751
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1710
- Ceres seq_id 1570752
- Location of start within SEQ ID NO 1709: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1711
- Ceres seq_id 1570753
- Location of start within SEQ ID NO 1709: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 776
- gi No. 3935169
- Description: (AC004557) F17L21.12 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 1711: from 1 to 68

Maximum Length Sequence:
related to:

Clone IDs:
27462
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1712
- Ceres seq_id 1570759
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1713
- Ceres seq_id 1570760
- Location of start within SEQ ID NO 1712: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1713: from 31 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 777
- gi No. 4033468
- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP40 >gi|2582641|emb|CAA67800| (X99437) splicing factor [Arabidopsis thaliana] >gi|2980800|emb|CAA18176.1| (AL022197) splicing factor At-SRp40 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 1713: from 28 to 200

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1714
- Ceres seq_id 1570761
- Location of start within SEQ ID NO 1712: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1714: from 4 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 778
- gi No. 4033468
- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP40 >gi|2582641|emb|CAA67800| (X99437) splicing factor [Arabidopsis thaliana] >gi|2980800|emb|CAA18176.1| (AL022197) splicing factor At-SRp40 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 1714: from 1 to 173

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1715
- Ceres seq_id 1570762
- Location of start within SEQ ID NO 1712: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1715: from 67 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 779
- gi No. 4033468

- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP40 >gi|2582641|emb|CAA67800| (X99437) splicing factor [Arabidopsis thaliana] >gi|2980800|emb|CAA18176.1| (AL022197) splicing factor At-SRp40 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 1715: from 1 to 141

Maximum Length Sequence:
related to:
Clone IDs:
28572
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1716
- Ceres seq_id 1570775

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1717
- Ceres seq_id 1570776
- Location of start within SEQ ID NO 1716: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1717: from 57 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 780
- gi No. 4725952
- Description: (AL049730) AIR1A-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 1717: from 24 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1718
- Ceres seq_id 1570777
- Location of start within SEQ ID NO 1716: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1718: from 34 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 781
- gi No. 4725952
- Description: (AL049730) AIR1A-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 1718: from 1 to 117

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1719
- Ceres seq_id 1570778
- Location of start within SEQ ID NO 1716: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1719: from 32 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 782
- gi No. 4725952
- Description: (AL049730) AIR1A-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 117
- Location of Alignment in SEQ ID NO 1719: from 1 to 115

Maximum Length Sequence:
related to:
Clone IDs:
28847
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1720
- Ceres seq_id 1570779
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1721
- Ceres seq_id 1570780
- Location of start within SEQ ID NO 1720: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HIT family
- Location within SEQ ID NO 1721: from 36 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 783
- gi No. 629858
- Description: protein kinase C inhibitor - maize
- % Identity: 82.2
- Alignment Length: 129
- Location of Alignment in SEQ ID NO 1721: from 19 to 147

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1722
- Ceres seq_id 1570781
- Location of start within SEQ ID NO 1720: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HIT family
- Location within SEQ ID NO 1722: from 18 to 121 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 784
- gi No. 629858
- Description: protein kinase C inhibitor - maize
- % Identity: 82.2
- Alignment Length: 129
- Location of Alignment in SEQ ID NO 1722: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
29009
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1723
- Ceres seq_id 1570782
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1724
- Ceres seq_id 1570783

- Location of start within SEQ ID NO 1723: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 785
- gi No. 2832664
- Description: (AL021710) pollen-specific protein - like [Arabidopsis thaliana]
- % Identity: 96.8
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 1724: from 20 to 143

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1725
- Ceres seq_id 1570784
- Location of start within SEQ ID NO 1723: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 786
- gi No. 2832664
- Description: (AL021710) pollen-specific protein - like [Arabidopsis thaliana]
- % Identity: 96.8
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 1725: from 1 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1726
- Ceres seq_id 1570785
- Location of start within SEQ ID NO 1723: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 787
- gi No. 2832664
- Description: (AL021710) pollen-specific protein - like [Arabidopsis thaliana]
- % Identity: 96.8
- Alignment Length: 124
- Location of Alignment in SEQ ID NO 1726: from 1 to 111

Maximum Length Sequence:
related to:
Clone IDs:
29227
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1727
- Ceres seq_id 1570786
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1728
- Ceres seq_id 1570787
- Location of start within SEQ ID NO 1727: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 1728: from 26 to 179 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 788
- gi No. 5454190
- Description: (AC005698) T3P18.4 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 1728: from 32 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1729
- Ceres seq_id 1570788
- Location of start within SEQ ID NO 1727: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 1729: from 1 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 789
- gi No. 5454190
- Description: (AC005698) T3P18.4 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 1729: from 1 to 152

Maximum Length Sequence:
related to:
Clone IDs:
29713
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1730
- Ceres seq_id 1570793

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1731
- Ceres seq_id 1570794
- Location of start within SEQ ID NO 1730: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 790
- gi No. 2129826
- Description: dynamin-like protein phragmoplastin 5 - soybean >gi|1218004 (U36430) SDL5A [Glycine max]
- % Identity: 91.3
- Alignment Length: 46
- Location of Alignment in SEQ ID NO 1731: from 1 to 46

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1732
- Ceres seq_id 1570795
- Location of start within SEQ ID NO 1730: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 791
- gi No. 2129826

- Description: dynamin-like protein phragmoplastin 5 - soybean >gi|1218004 (U36430) SDL5A [Glycine max]
- % Identity: 91.3
- Alignment Length: 46
- Location of Alignment in SEQ ID NO 1732: from 1 to 24

Maximum Length Sequence:
related to:
Clone IDs:
29893
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1733
- Ceres seq_id 1570796
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1734
- Ceres seq_id 1570797
- Location of start within SEQ ID NO 1733: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA polymerases N / 8 kDa subunit
- Location within SEQ ID NO 1734: from 1 to 60 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 792
- gi No. 2833375
- Description: DNA-DIRECTED RNA POLYMERASE II 8.2 KD POLYPEPTIDE (RPB10) (RP10) (ABC10) >gi|533690 (U12133) RNA polymerase II subunit RPB10 homolog; similar to yeast RNA polymerase II subunit RPB10, Swiss-Prot Accession Number P22139 [Brassica napus]
- % Identity: 89.9
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 1734: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
42589
29919
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1735
- Ceres seq_id 1570798
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1736
- Ceres seq_id 1570799
- Location of start within SEQ ID NO 1735: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S16
- Location within SEQ ID NO 1736: from 30 to 91 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1737
- Ceres seq_id 1570800
- Location of start within SEQ ID NO 1735: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Ribosomal protein S16
- Location within SEQ ID NO 1737: from 9 to 70 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
30105
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1738
- Ceres seq_id 1570805
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1739
- Ceres seq_id 1570806
- Location of start within SEQ ID NO 1738: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial RNA polymerase, alpha chain
- Location within SEQ ID NO 1739: from 3 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 793
- gi No. 133409
- Description: DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN >gi|66962|pir||RNNTA DNA-directed RNA polymerase (EC 2.7.7.6) alpha chain - common tobacco chloroplast >gi|11860|emb|CAA77376| (Z00044) RNA polymerase alpha subunit [Nicotiana tabacum]
- % Identity: 76.8
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 1739: from 3 to 157

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1740
- Ceres seq_id 1570807
- Location of start within SEQ ID NO 1738: at 214 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial RNA polymerase, alpha chain
- Location within SEQ ID NO 1740: from 1 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 794
- gi No. 133409
- Description: DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN >gi|66962|pir||RNNTA DNA-directed RNA polymerase (EC 2.7.7.6) alpha chain - common tobacco chloroplast >gi|11860|emb|CAA77376| (Z00044) RNA polymerase alpha subunit [Nicotiana tabacum]
- % Identity: 76.8
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 1740: from 1 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1741
- Ceres seq_id 1570808
- Location of start within SEQ ID NO 1738: at 262 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Bacterial RNA polymerase, alpha chain
- Location within SEQ ID NO 1741: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 795
- gi No. 133409
- Description: DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN >gi|66962|pir||RNNTA DNA-directed RNA polymerase (EC 2.7.7.6) alpha chain - common tobacco chloroplast >gi|11860|emb|CAA77376| (Z00044) RNA polymerase alpha subunit [Nicotiana tabacum]
- % Identity: 76.8
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 1741: from 1 to 81

Maximum Length Sequence:
related to:
Clone IDs:
30167
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1742
- Ceres seq_id 1570816
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1743
- Ceres seq_id 1570817
- Location of start within SEQ ID NO 1742: at 41 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 796
- gi No. 1780757
- Description: (Y10291) highly expressed in proliferating cells [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 1743: from 1 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1744
- Ceres seq_id 1570818
- Location of start within SEQ ID NO 1742: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 797
- gi No. 1780757
- Description: (Y10291) highly expressed in proliferating cells [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 1744: from 1 to 73

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1745
- Ceres seq_id 1570819
- Location of start within SEQ ID NO 1742: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 798
- gi No. 1780757
- Description: (Y10291) highly expressed in proliferating cells [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 1745: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
30187
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1746
- Ceres seq_id 1570823
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1747
- Ceres seq_id 1570824
- Location of start within SEQ ID NO 1746: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1747: from 45 to 72 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1748
- Ceres seq_id 1570825
- Location of start within SEQ ID NO 1746: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1748: from 29 to 56 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1749
- Ceres seq_id 1570826
- Location of start within SEQ ID NO 1746: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 1749: from 25 to 52 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
30530
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1750
- Ceres seq_id 1570827
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1751
- Ceres seq_id 1570828

- Location of start within SEQ ID NO 1750: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1751: from 7 to 73 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
30903
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1752
- Ceres seq_id 1570829
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1753
- Ceres seq_id 1570830
- Location of start within SEQ ID NO 1752: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L36e
- Location within SEQ ID NO 1753: from 28 to 124 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1754
- Ceres seq_id 1570831
- Location of start within SEQ ID NO 1752: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L36e
- Location within SEQ ID NO 1754: from 2 to 98 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
31676
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1755
- Ceres seq_id 1570836
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1756
- Ceres seq_id 1570837
- Location of start within SEQ ID NO 1755: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0001
- Location within SEQ ID NO 1756: from 30 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1757

- Ceres seq_id 1570838
- Location of start within SEQ ID NO 1755: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0001
- Location within SEQ ID NO 1757: from 25 to 109 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1758
- Ceres seq_id 1570839
- Location of start within SEQ ID NO 1755: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0001
- Location within SEQ ID NO 1758: from 10 to 94 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
32371
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1759
- Ceres seq_id 1570852

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1760
- Ceres seq_id 1570853
- Location of start within SEQ ID NO 1759: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 1760: from 1 to 182 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 799
- gi No. 3915033
- Description: ACYL-[ACYL-CARRIER PROTEIN] DESATURASE PRECURSOR (STEAROYL-ACP DESATURASE) >gi|1354933 (U58141) stearoyl-ACP desaturase [Olea europaea]
- % Identity: 73.1
- Alignment Length: 186
- Location of Alignment in SEQ ID NO 1760: from 1 to 186

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1761
- Ceres seq_id 1570854
- Location of start within SEQ ID NO 1759: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 1761: from 1 to 131 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 800

- gi No. 3915033
- Description: ACYL-[ACYL-CARRIER PROTEIN] DESATURASE PRECURSOR (STEAROYL-ACP DESATURASE) >gi|1354933 (U58141) stearoyl-ACP desaturase [Olea europaea]
- % Identity: 73.1
- Alignment Length: 186
- Location of Alignment in SEQ ID NO 1761: from 1 to 135

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1762
- Ceres seq_id 1570855
- Location of start within SEQ ID NO 1759: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 1762: from 1 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 801
- gi No. 3915033
- Description: ACYL-[ACYL-CARRIER PROTEIN] DESATURASE PRECURSOR (STEAROYL-ACP DESATURASE) >gi|1354933 (U58141) stearoyl-ACP desaturase [Olea europaea]
- % Identity: 73.1
- Alignment Length: 186
- Location of Alignment in SEQ ID NO 1762: from 1 to 111

Maximum Length Sequence:
related to:
Clone IDs:
32457
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1763
- Ceres seq_id 1570862

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1764
- Ceres seq_id 1570863
- Location of start within SEQ ID NO 1763: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1764: from 27 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 802
- gi No. 4759158
- Description: ref|NP_004588.1|pSNRPD2| small nuclear ribonucleoprotein D2 polypeptide (16.5kD); SNRNP CORE PROTEIN D2; SM-D2 >gi|1173455|sp|P43330|SMD2_HUMAN SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 (SNRNP CORE PROTEIN D2) (SM-D2) >gi|2136168|pir||I38861 small
- % Identity: 80
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1764: from 1 to 108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1765
- Ceres seq_id 1570864
- Location of start within SEQ ID NO 1763: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1765: from 23 to 102 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 803
- gi No. 4759158
- Description: ref|NP_004588.1|pSNRPD2| small nuclear ribonucleoprotein D2 polypeptide (16.5kD); SNRNP CORE PROTEIN D2; SM-D2 >gi|1173455|sp|P43330|SMD2_HUMAN SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 (SNRNP CORE PROTEIN D2) (SM-D2) >gi|2136168|pir||I38861 small
- % Identity: 80
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1765: from 1 to 104

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1766
- Ceres seq_id 1570865
- Location of start within SEQ ID NO 1763: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1766: from 1 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 804
- gi No. 4759158
- Description: ref|NP_004588.1|pSNRPD2| small nuclear ribonucleoprotein D2 polypeptide (16.5kD); SNRNP CORE PROTEIN D2; SM-D2 >gi|1173455|sp|P43330|SMD2_HUMAN SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 (SNRNP CORE PROTEIN D2) (SM-D2) >gi|2136168|pir||I38861 small
- % Identity: 80
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 1766: from 1 to 81

Maximum Length Sequence:
related to:
Clone IDs:
32642
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1767
- Ceres seq_id 1570869

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1768
- Ceres seq_id 1570870
- Location of start within SEQ ID NO 1767: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 1768: from 15 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 805
- gi No. 122106
- Description: HISTONE H4 >gi|70771|pir||HSZM4 histone H4 - maize >gi|81642|pir||S06904 histone H4 - Arabidopsis thaliana >gi|2119028|pir||S60475 histone H4 - garden pea >gi|21795|emb|CAA24924| (X00043) histone H4 [Triticum aestivum] [Arabidopsis thaliana]

- % Identity: 100
- Alignment Length: 103
- Location of Alignment in SEQ ID NO 1768: from 14 to 116

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1769
- Ceres seq_id 1570871
- Location of start within SEQ ID NO 1767: at 42 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 1769: from 2 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 806
- gi No. 122106
- Description: HISTONE H4 >gi|70771|pir||HSZM4 histone H4 - maize >gi|81642|pir||S06904 histone H4 - Arabidopsis thaliana >gi|2119028|pir||S60475 histone H4 - garden pea >gi|21795|emb|CAA24924|(X00043) histone H4 [Triticum aestivum] [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 103
- Location of Alignment in SEQ ID NO 1769: from 1 to 103

Maximum Length Sequence:
related to:
Clone IDs:
34079
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1770
- Ceres seq_id 1570895

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1771
- Ceres seq_id 1570896
- Location of start within SEQ ID NO 1770: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FYVE zinc finger
- Location within SEQ ID NO 1771: from 118 to 183 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1772
- Ceres seq_id 1570897
- Location of start within SEQ ID NO 1770: at 31 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FYVE zinc finger
- Location within SEQ ID NO 1772: from 108 to 173 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1773
- Ceres seq_id 1570898
- Location of start within SEQ ID NO 1770: at 220 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FYVE zinc finger
- Location within SEQ ID NO 1773: from 45 to 110 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34741
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1774
- Ceres seq_id 1570913
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1775
- Ceres seq_id 1570914
- Location of start within SEQ ID NO 1774: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 807
- gi No. 3063442
- Description: (AC003981) F22013.6 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 1775: from 2 to 124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1776
- Ceres seq_id 1570915
- Location of start within SEQ ID NO 1774: at 4 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 808
- gi No. 3063442
- Description: (AC003981) F22013.6 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 1776: from 1 to 123

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1777
- Ceres seq_id 1570916
- Location of start within SEQ ID NO 1774: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 809
- gi No. 3063442
- Description: (AC003981) F22013.6 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 1777: from 1 to 65

Maximum Length Sequence:
related to:

Clone IDs:
35090
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1778
- Ceres seq_id 1570931
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1779
- Ceres seq_id 1570932
- Location of start within SEQ ID NO 1778: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1779: from 55 to 121 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
35505
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1780
- Ceres seq_id 1570936
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1781
- Ceres seq_id 1570937
- Location of start within SEQ ID NO 1780: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial mutT protein
- Location within SEQ ID NO 1781: from 23 to 62 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1782
- Ceres seq_id 1570938
- Location of start within SEQ ID NO 1780: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1783
- Ceres seq_id 1570939
- Location of start within SEQ ID NO 1780: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
35952
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1784
- Ceres seq_id 1570951

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1785
- Ceres seq_id 1570952
- Location of start within SEQ ID NO 1784: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glyoxalase
- Location within SEQ ID NO 1785: from 35 to 156 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1786
- Ceres seq_id 1570953
- Location of start within SEQ ID NO 1784: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glyoxalase
- Location within SEQ ID NO 1786: from 13 to 134 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1787
- Ceres seq_id 1570954
- Location of start within SEQ ID NO 1784: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glyoxalase
- Location within SEQ ID NO 1787: from 1 to 84 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37986
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1788
- Ceres seq_id 1570974

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1789
- Ceres seq_id 1570975
- Location of start within SEQ ID NO 1788: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ER lumen protein retaining receptor
- Location within SEQ ID NO 1789: from 67 to 277 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1790
- Ceres seq_id 1570976
- Location of start within SEQ ID NO 1788: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ER lumen protein retaining receptor
- Location within SEQ ID NO 1790: from 48 to 258 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1791
- Ceres seq_id 1570977
- Location of start within SEQ ID NO 1788: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ER lumen protein retaining receptor
- Location within SEQ ID NO 1791: from 16 to 226 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
38794
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1792
- Ceres seq_id 1570998
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1793
- Ceres seq_id 1570999
- Location of start within SEQ ID NO 1792: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1793: from 90 to 160 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 810
- gi No. 1171978
- Description: POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1793: from 1 to 502

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1794
- Ceres seq_id 1571000
- Location of start within SEQ ID NO 1792: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1794: from 31 to 101 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 811
- gi No. 1171978

- Description: POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1794: from 1 to 443

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1795
- Ceres seq_id 1571001
- Location of start within SEQ ID NO 1792: at 359 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Poly-adenylate binding protein, unique domain.
- Location within SEQ ID NO 1795: from 294 to 365 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 812
- gi No. 1171978
- Description: POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 502
- Location of Alignment in SEQ ID NO 1795: from 1 to 383

Maximum Length Sequence:
related to:
Clone IDs:
39583
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1796
- Ceres seq_id 1571014
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1797
- Ceres seq_id 1571015
- Location of start within SEQ ID NO 1796: at 33 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1798
- Ceres seq_id 1571016
- Location of start within SEQ ID NO 1796: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1799
- Ceres seq_id 1571017
- Location of start within SEQ ID NO 1796: at 496 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 1799: from 1 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 813
- gi No. 2119846
- Description: chlorophyll a/b-binding protein type I precursor Lhb1B2 - Arabidopsis thaliana >gi|16364|emb|CAA45790| (X64460) photosystem II type I chlorophyll a /b binding protein [Arabidopsis thaliana] thaliana]
- % Identity: 98.6
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 1799: from 1 to 99

Maximum Length Sequence:
related to:
Clone IDs:
39877
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1800
- Ceres seq_id 1571021
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1801
- Ceres seq_id 1571022
- Location of start within SEQ ID NO 1800: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 1801: from 88 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 814
- gi No. 3868857
- Description: (AB013886) RAV1 [Arabidopsis thaliana]
- % Identity: 73.6
- Alignment Length: 345
- Location of Alignment in SEQ ID NO 1801: from 34 to 361

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1802
- Ceres seq_id 1571023
- Location of start within SEQ ID NO 1800: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 1802: from 59 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 815
- gi No. 3868857
- Description: (AB013886) RAV1 [Arabidopsis thaliana]
- % Identity: 73.6
- Alignment Length: 345
- Location of Alignment in SEQ ID NO 1802: from 5 to 332

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1803
- Ceres seq_id 1571024
- Location of start within SEQ ID NO 1800: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 1803: from 56 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 816
- gi No. 3868857
- Description: (AB013886) RAV1 [Arabidopsis thaliana]
- % Identity: 73.6
- Alignment Length: 345
- Location of Alignment in SEQ ID NO 1803: from 2 to 329

Maximum Length Sequence:
related to:
Clone IDs:
40272
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1804
- Ceres seq_id 1571031
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1805
- Ceres seq_id 1571032
- Location of start within SEQ ID NO 1804: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1806
- Ceres seq_id 1571033
- Location of start within SEQ ID NO 1804: at 249 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1807
- Ceres seq_id 1571034
- Location of start within SEQ ID NO 1804: at 643 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0016
- Location within SEQ ID NO 1807: from 1 to 112 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
41988
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1808
- Ceres seq_id 1571042
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1809
- Ceres seq_id 1571043
- Location of start within SEQ ID NO 1808: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 817
- gi No. 2129641
- Description: major latex protein type 1 - Arabidopsis thaliana >gi|1107493|emb|CAA63026| (X91960) major latex protein type1 [Arabidopsis thaliana]
- % Identity: 71.3
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 1809: from 1 to 150

Maximum Length Sequence:
related to:
Clone IDs:
42276
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1810
- Ceres seq_id 1571044
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1811
- Ceres seq_id 1571045
- Location of start within SEQ ID NO 1810: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1812
- Ceres seq_id 1571046
- Location of start within SEQ ID NO 1810: at 27 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 818
- gi No. 2677830
- Description: (U93168) ribosomal protein L12 [Prunus armeniaca]
- % Identity: 90.5
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1812: from 1 to 21

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1813
- Ceres seq_id 1571047
- Location of start within SEQ ID NO 1810: at 376 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L11
- Location within SEQ ID NO 1813: from 1 to 29 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 819
- gi No. 2677830
- Description: (U93168) ribosomal protein L12 [Prunus armeniaca]
- % Identity: 87.4
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 1813: from 1 to 48

Maximum Length Sequence:
related to:
Clone IDs:
5171
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1814
- Ceres seq_id 1571062
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1815
- Ceres seq_id 1571063
- Location of start within SEQ ID NO 1814: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1815: from 43 to 102 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 820
- gi No. 3157948
- Description: (AC002131) Contains similarity to thyroidal receptor for N-acetylglucosamine gb|A27284 from Homo sapiens. EST gb|N65493 comes from this gene. [Arabidopsis thaliana]
- % Identity: 84.3
- Alignment Length: 108
- Location of Alignment in SEQ ID NO 1815: from 12 to 104

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1816
- Ceres seq_id 1571064
- Location of start within SEQ ID NO 1814: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1816: from 32 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 821
- gi No. 3157948
- Description: (AC002131) Contains similarity to thyroidal receptor for N-acetylglucosamine gb|A27284 from Homo sapiens. EST gb|N65493 comes from this gene. [Arabidopsis thaliana]
- % Identity: 84.3
- Alignment Length: 108
- Location of Alignment in SEQ ID NO 1816: from 1 to 93

Maximum Length Sequence:
related to:
Clone IDs:
16143
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1817
- Ceres seq_id 1571079
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1818
- Ceres seq_id 1571080
- Location of start within SEQ ID NO 1817: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pathogenesis-related protein Bet v I family
- Location within SEQ ID NO 1818: from 5 to 155 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6256
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1819
- Ceres seq_id 1571085
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1820
- Ceres seq_id 1571086
- Location of start within SEQ ID NO 1819: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1820: from 36 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 822
- gi No. 4768968
- Description: (AF140219) nuclear cap-binding protein; CBP20 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 138
- Location of Alignment in SEQ ID NO 1820: from 1 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1821
- Ceres seq_id 1571087
- Location of start within SEQ ID NO 1819: at 272 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 823
- gi No. 4768968
- Description: (AF140219) nuclear cap-binding protein; CBP20 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 138
- Location of Alignment in SEQ ID NO 1821: from 1 to 74

Maximum Length Sequence:
related to:
Clone IDs:
6333
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1822
- Ceres seq_id 1571092
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1823
- Ceres seq_id 1571093
- Location of start within SEQ ID NO 1822: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 824
- gi No. 464258
- Description: NADH-UBIQUINONE OXIDOREDUCTASE 22.5 KD SUBUNIT (COMPLEX I-22.5KD) (CI-22.5KD) >gi|539041|pir||F49732 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 22.5K chain - potato mitochondrion (fragment)
- % Identity: 81.1
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 1823: from 36 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1824
- Ceres seq_id 1571094
- Location of start within SEQ ID NO 1822: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1825
- Ceres seq_id 1571095
- Location of start within SEQ ID NO 1822: at 299 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6906
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1826
- Ceres seq_id 1571116
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1827
- Ceres seq_id 1571117
- Location of start within SEQ ID NO 1826: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 825
- gi No. 3335335
- Description: (AC004512) ESTs gb|F14113 and gb|T42122 come from this region. [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 1827: from 39 to 330

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1828
- Ceres seq_id 1571118
- Location of start within SEQ ID NO 1826: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 826
- gi No. 3335335
- Description: (AC004512) ESTs gb|F14113 and gb|T42122 come from this region. [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 1828: from 1 to 292

Maximum Length Sequence:
related to:
Clone IDs:
7015
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1829
- Ceres seq_id 1571119
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1830
- Ceres seq_id 1571120
- Location of start within SEQ ID NO 1829: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 827
- gi No. 1363489
- Description: thioglucosidase (EC 3.2.3.1) 3D precursor - Arabidopsis thaliana >gi|984052|emb|CAA61592| (X89413) thioglucoside glucohydrolase [Arabidopsis thaliana] >gi|5524767|emb|CAB50792.1| (AJ243490) thioglucoside glucohydrolase [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 1830: from 6 to 42

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1831
- Ceres seq_id 1571121
- Location of start within SEQ ID NO 1829: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 828
- gi No. 1363489
- Description: thioglucosidase (EC 3.2.3.1) 3D precursor - Arabidopsis thaliana >gi|984052|emb|CAA61592| (X89413) thioglucoside glucohydrolase [Arabidopsis thaliana] >gi|5524767|emb|CAB50792.1| (AJ243490) thioglucoside glucohydrolase [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 1831: from 1 to 37

Maximum Length Sequence:
related to:
Clone IDs:
7182
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1832
- Ceres seq_id 1571126
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1833
- Ceres seq_id 1571127

- Location of start within SEQ ID NO 1832: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L34e
- Location within SEQ ID NO 1833: from 21 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 829
- gi No. 2500376
- Description: 60S RIBOSOMAL PROTEIN L34 >gi|4262177|gb|AAD14494|(AC005508) 23552 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 1833: from 17 to 136

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1834
- Ceres seq_id 1571128
- Location of start within SEQ ID NO 1832: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1835
- Ceres seq_id 1571129
- Location of start within SEQ ID NO 1832: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L34e
- Location within SEQ ID NO 1835: from 5 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 830
- gi No. 2500376
- Description: 60S RIBOSOMAL PROTEIN L34 >gi|4262177|gb|AAD14494|(AC005508) 23552 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 1835: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
7420
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1836
- Ceres seq_id 1571134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1837
- Ceres seq_id 1571135
- Location of start within SEQ ID NO 1836: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1837: from 10 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 831
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 99.4
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 1837: from 1 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1838
- Ceres seq_id 1571136
- Location of start within SEQ ID NO 1836: at 241 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1838: from 30 to 111 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 832
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 99.4
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 1838: from 1 to 114

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1839
- Ceres seq_id 1571137
- Location of start within SEQ ID NO 1836: at 262 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 1839: from 23 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 833
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 99.4
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 1839: from 1 to 107

Maximum Length Sequence:
related to:
Clone IDs:
7767
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1840
- Ceres seq_id 1571138
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1841
- Ceres seq_id 1571139
- Location of start within SEQ ID NO 1840: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 834
- gi No. 3766248
- Description: (Y18227) blue copper binding-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 196
- Location of Alignment in SEQ ID NO 1841: from 1 to 196

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1842
- Ceres seq_id 1571140
- Location of start within SEQ ID NO 1840: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 835
- gi No. 3766248
- Description: (Y18227) blue copper binding-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 196
- Location of Alignment in SEQ ID NO 1842: from 1 to 160

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1843
- Ceres seq_id 1571141
- Location of start within SEQ ID NO 1840: at 382 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 836
- gi No. 3766248
- Description: (Y18227) blue copper binding-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 196
- Location of Alignment in SEQ ID NO 1843: from 1 to 109

Maximum Length Sequence:
related to:
Clone IDs:
8355

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1844
- Ceres seq_id 1571153

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1845
- Ceres seq_id 1571154
- Location of start within SEQ ID NO 1844: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Heme-binding domain in cytochrome b5 and oxidoreductases
- Location within SEQ ID NO 1845: from 74 to 147 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1846
- Ceres seq_id 1571155
- Location of start within SEQ ID NO 1844: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Heme-binding domain in cytochrome b5 and oxidoreductases
- Location within SEQ ID NO 1846: from 48 to 121 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
857
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1847
- Ceres seq_id 1571159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1848
- Ceres seq_id 1571160
- Location of start within SEQ ID NO 1847: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 837
- gi No. 2982253
- Description: (AF051209) CROC-1-like protein [Picea mariana]
- % Identity: 81.3
- Alignment Length: 16
- Location of Alignment in SEQ ID NO 1848: from 3 to 18

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1849
- Ceres seq_id 1571161
- Location of start within SEQ ID NO 1847: at 12 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 838
- gi No. 2982253
- Description: (AF051209) CROC-1-like protein [Picea mariana]
- % Identity: 81.3
- Alignment Length: 16
- Location of Alignment in SEQ ID NO 1849: from 1 to 15

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1850
- Ceres seq_id 1571162
- Location of start within SEQ ID NO 1847: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
8586
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1851
- Ceres seq_id 1571163
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1852
- Ceres seq_id 1571164
- Location of start within SEQ ID NO 1851: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 839
- gi No. 4914432
- Description: (AL050351) ribosomal protein S25 [Arabidopsis thaliana]
- % Identity: 95.3
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 1852: from 1 to 107

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1853
- Ceres seq_id 1571165
- Location of start within SEQ ID NO 1851: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 840
- gi No. 4914432
- Description: (AL050351) ribosomal protein S25 [Arabidopsis thaliana]
- % Identity: 95.3
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 1853: from 1 to 68

Maximum Length Sequence:
related to:
Clone IDs:
8841
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1854
- Ceres seq_id 1571166
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1855
- Ceres seq_id 1571167
- Location of start within SEQ ID NO 1854: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 841
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]
- % Identity: 74.3

- Alignment Length: 152
- Location of Alignment in SEQ ID NO 1855: from 25 to 176

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1856
- Ceres seq_id 1571168
- Location of start within SEQ ID NO 1854: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 842
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]
- % Identity: 74.3
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 1856: from 1 to 152

Maximum Length Sequence:
related to:
Clone IDs:
17402
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1857
- Ceres seq_id 1571206

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1858
- Ceres seq_id 1571207
- Location of start within SEQ ID NO 1857: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 1858: from 28 to 68 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
18060
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1859
- Ceres seq_id 1571220

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1860
- Ceres seq_id 1571221
- Location of start within SEQ ID NO 1859: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 1860: from 5 to 53 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 843
- gi No. 3850571
- Description: (AC005278) Similar to gb|U85207 snRNP core Sm protein homolog Sm-X5 from Mus musculus. EST gb|AA612141 comes from this gene. [Arabidopsis thaliana]

- % Identity: 89.1
- Alignment Length: 55
- Location of Alignment in SEQ ID NO 1860: from 2 to 56

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1861
- Ceres seq_id 1571222
- Location of start within SEQ ID NO 1859: at 244 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 844
- gi No. 2463648
- Description: (U85207) snRNP core Sm protein homolog Sm-X5 [Mus musculus] >gi|4529895|gb|AAD21818.1| (AF134726) snRNP [Homo sapiens] >gi|5701852|emb|CAB52190.1| (AJ245416) G7b protein [Homo sapiens]
- % Identity: 80.5
- Alignment Length: 41
- Location of Alignment in SEQ ID NO 1861: from 1 to 37

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1862
- Ceres seq_id 1571223
- Location of start within SEQ ID NO 1859: at 308 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
18140

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1863
- Ceres seq_id 1571228

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1864
- Ceres seq_id 1571229
- Location of start within SEQ ID NO 1863: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KH domain
- Location within SEQ ID NO 1864: from 153 to 202 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1865
- Ceres seq_id 1571230
- Location of start within SEQ ID NO 1863: at 312 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KH domain
- Location within SEQ ID NO 1865: from 99 to 148 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1866
- Ceres seq_id 1571231
- Location of start within SEQ ID NO 1863: at 824 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20407
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1867
- Ceres seq_id 1571235
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1868
- Ceres seq_id 1571236
- Location of start within SEQ ID NO 1867: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 1868: from 26 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 845
- gi No. 2290392
- Description: (U90353) IgG and IgE immunoreactive antigen recognized by sera from patients with strongyloidiasis [Strongyloides stercoralis]
- % Identity: 77.6
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 1868: from 52 to 99

Maximum Length Sequence:
related to:
Clone IDs:
21903
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1869
- Ceres seq_id 1571241
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1870
- Ceres seq_id 1571242
- Location of start within SEQ ID NO 1869: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectinesterase
- Location within SEQ ID NO 1870: from 1 to 215 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22146
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1871

- Ceres seq_id 1571243

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1872
- Ceres seq_id 1571244
- Location of start within SEQ ID NO 1871: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 846
- gi No. 122106
- Description: HISTONE H4 >gi|70771|pir||HSZM4 histone H4 - maize >gi|81642|pir||S06904 histone H4 - Arabidopsis thaliana >gi|2119028|pir||S60475 histone H4 - garden pea >gi|21795|emb|CAA24924|(X00043) histone H4 [Triticum aestivum] [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 1872: from 1 to 49

Maximum Length Sequence:
related to:
Clone IDs:
22430

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1873
- Ceres seq_id 1571245

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1874
- Ceres seq_id 1571246
- Location of start within SEQ ID NO 1873: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TBC domain
- Location within SEQ ID NO 1874: from 381 to 463 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1875
- Ceres seq_id 1571247
- Location of start within SEQ ID NO 1873: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TBC domain
- Location within SEQ ID NO 1875: from 372 to 454 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1876
- Ceres seq_id 1571248
- Location of start within SEQ ID NO 1873: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TBC domain
- Location within SEQ ID NO 1876: from 370 to 452 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22697
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1877
- Ceres seq_id 1571257
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1878
- Ceres seq_id 1571258
- Location of start within SEQ ID NO 1877: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 1878: from 25 to 121 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1879
- Ceres seq_id 1571259
- Location of start within SEQ ID NO 1877: at 26 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 1879: from 17 to 113 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1880
- Ceres seq_id 1571260
- Location of start within SEQ ID NO 1877: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L17
- Location within SEQ ID NO 1880: from 3 to 99 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22551
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1881
- Ceres seq_id 1571261
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1882
- Ceres seq_id 1571262
- Location of start within SEQ ID NO 1881: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Poly-adenylate binding protein, unique domain.
- Location within SEQ ID NO 1882: from 12 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 847
- gi No. 1171978
- Description: POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [Arabidopsis thaliana]
- % Identity: 70.8
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 1882: from 2 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1883
- Ceres seq_id 1571264
- Location of start within SEQ ID NO 1881: at 10 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Poly-adenylate binding protein, unique domain.
- Location within SEQ ID NO 1883: from 9 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 848
- gi No. 1171978
- Description: POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [Arabidopsis thaliana]
- % Identity: 70.8
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 1883: from 1 to 83

Maximum Length Sequence:
related to:
Clone IDs:
2403
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1884
- Ceres seq_id 1571265
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1885
- Ceres seq_id 1571266
- Location of start within SEQ ID NO 1884: at 282 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 1885: from 2 to 75 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 849
- gi No. 1362009
- Description: ubiquitin-like protein 7 - Arabidopsis thaliana
- % Identity: 99.4
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 1885: from 1 to 77

Maximum Length Sequence:
related to:
Clone IDs:
2443

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1886
- Ceres seq_id 1571271

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1887
- Ceres seq_id 1571272
- Location of start within SEQ ID NO 1886: at 207 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Synaptobrevin
- Location within SEQ ID NO 1887: from 118 to 209 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 850
- gi No. 2924792
- Description: (AC002334) similar to synaptobrevin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 1887: from 1 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1888
- Ceres seq_id 1571273
- Location of start within SEQ ID NO 1886: at 441 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Synaptobrevin
- Location within SEQ ID NO 1888: from 40 to 131 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 851
- gi No. 2924792
- Description: (AC002334) similar to synaptobrevin [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 1888: from 1 to 143

Maximum Length Sequence:
related to:
Clone IDs:
25590

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1889
- Ceres seq_id 1571291

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1890
- Ceres seq_id 1571292
- Location of start within SEQ ID NO 1889: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1890: from 126 to 249 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 852
- gi No. 1419388

- Description: (X98925) stromal ascorbate peroxidase [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 351
- Location of Alignment in SEQ ID NO 1890: from 4 to 353

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1891
- Ceres seq_id 1571293
- Location of start within SEQ ID NO 1889: at 19 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 1891: from 120 to 243 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 853
- gi No. 1419388
- Description: (X98925) stromal ascorbate peroxidase [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 351
- Location of Alignment in SEQ ID NO 1891: from 1 to 347

Maximum Length Sequence:
related to:
Clone IDs:
27210
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1892
- Ceres seq_id 1571304
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1893
- Ceres seq_id 1571305
- Location of start within SEQ ID NO 1892: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Calreticulin family
- Location within SEQ ID NO 1893: from 24 to 336 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 854
- gi No. 3249100
- Description: (AC003114) Match to calreticulin (AtCRTL) mRNA gb|U27698 and DNA gb|U66344. ESTs gb|T45719, gb|T22451, gb|H36323 and gb|AA042519 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 430
- Location of Alignment in SEQ ID NO 1893: from 1 to 421

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1894
- Ceres seq_id 1571306
- Location of start within SEQ ID NO 1892: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Calreticulin family
- Location within SEQ ID NO 1894: from 21 to 333 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 855
- gi No. 3249100
- Description: (AC003114) Match to calreticulin (AtCRTL) mRNA gb|U27698 and DNA gb|U66344. ESTs gb|T45719, gb|T22451, gb|H36323 and gb|AA042519 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 430
- Location of Alignment in SEQ ID NO 1894: from 1 to 418

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1895
- Ceres seq_id 1571307
- Location of start within SEQ ID NO 1892: at 411 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Calreticulin family
- Location within SEQ ID NO 1895: from 1 to 224 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 856
- gi No. 3249100
- Description: (AC003114) Match to calreticulin (AtCRTL) mRNA gb|U27698 and DNA gb|U66344. ESTs gb|T45719, gb|T22451, gb|H36323 and gb|AA042519 come from this gene. [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 430
- Location of Alignment in SEQ ID NO 1895: from 1 to 309

Maximum Length Sequence:
related to:
Clone IDs:
27763
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1896
- Ceres seq_id 1571312

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1897
- Ceres seq_id 1571313
- Location of start within SEQ ID NO 1896: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1897: from 78 to 339 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 857
- gi No. 2190553
- Description: (AC001229) Strong similarity to Arabidopsis zeta-crystallin-like protein (gb|Z49268). [Arabidopsis thaliana]
- % Identity: 96.6
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 1897: from 78 to 339

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1898
- Ceres seq_id 1571314
- Location of start within SEQ ID NO 1896: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1898: from 52 to 313 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 858
- gi No. 2190553
- Description: (AC001229) Strong similarity to Arabidopsis zeta-crystallin-like protein (gb|Z49268). [Arabidopsis thaliana]
- % Identity: 96.6
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 1898: from 52 to 313

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1899
- Ceres seq_id 1571315
- Location of start within SEQ ID NO 1896: at 284 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 1899: from 20 to 281 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 859
- gi No. 2190553
- Description: (AC001229) Strong similarity to Arabidopsis zeta-crystallin-like protein (gb|Z49268). [Arabidopsis thaliana]
- % Identity: 96.6
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 1899: from 20 to 281

Maximum Length Sequence:
related to:
Clone IDs:
29774
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1900
- Ceres seq_id 1571333
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1901
- Ceres seq_id 1571334
- Location of start within SEQ ID NO 1900: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 860
- gi No. 2764524
- Description: (AJ000520) Rieske iron-sulfur protein Tic55 [Pisum sativum]
- % Identity: 70.1
- Alignment Length: 386
- Location of Alignment in SEQ ID NO 1901: from 13 to 382

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1902
- Ceres seq_id 1571335
- Location of start within SEQ ID NO 1900: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 861
- gi No. 2764524
- Description: (AJ000520) Rieske iron-sulfur protein Tic55 [Pisum sativum]
- % Identity: 70.1
- Alignment Length: 386
- Location of Alignment in SEQ ID NO 1902: from 1 to 370

Maximum Length Sequence:
related to:
Clone IDs:
30022

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1903
- Ceres seq_id 1571340

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1904
- Ceres seq_id 1571341
- Location of start within SEQ ID NO 1903: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1904: from 179 to 251 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 862
- gi No. 4835793
- Description: (AC007296) Similar to gb|U90212 DNA binding protein ACBF from Nicotiana tabacum and contains 3 PF|00076 RNA recognition motif domains. ESTs gb|T44278, gb|R65195, gb|N65904, gb|H37499, gb|R90487, gb|N95952, gb|T4427...
- % Identity: 99.5
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1904: from 23 to 427

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1905
- Ceres seq_id 1571342
- Location of start within SEQ ID NO 1903: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1905: from 157 to 229 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 863
- gi No. 4835793
- Description: (AC007296) Similar to gb|U90212 DNA binding protein ACBF from Nicotiana tabacum and contains 3 PF|00076 RNA recognition motif domains. ESTs gb|T44278, gb|R65195, gb|N65904, gb|H37499, gb|R90487, gb|N95952, gb|T4427...
- % Identity: 99.5
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1905: from 1 to 405

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1906
- Ceres seq_id 1571343
- Location of start within SEQ ID NO 1903: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1906: from 156 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 864
- gi No. 4835793
- Description: (AC007296) Similar to gb|U90212 DNA binding protein ACBF from Nicotiana tabacum and contains 3 PF|00076 RNA recognition motif domains. ESTs gb|T44278, gb|R65195, gb|N65904, gb|H37499, gb|R90487, gb|N95952, gb|T4427...
- % Identity: 99.5
- Alignment Length: 405
- Location of Alignment in SEQ ID NO 1906: from 1 to 404

Maximum Length Sequence:
related to:
Clone IDs:
30772
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1907
- Ceres seq_id 1571348
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1908
- Ceres seq_id 1571349
- Location of start within SEQ ID NO 1907: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 865
- gi No. 3090403
- Description: (AJ005671) cytochrome b6f complex subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 46
- Location of Alignment in SEQ ID NO 1908: from 56 to 101

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1909
- Ceres seq_id 1571350
- Location of start within SEQ ID NO 1907: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 866
- gi No. 3090403
- Description: (AJ005671) cytochrome b6f complex subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 46
- Location of Alignment in SEQ ID NO 1909: from 30 to 75

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1910
- Ceres seq_id 1571351
- Location of start within SEQ ID NO 1907: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 867
- gi No. 3090403
- Description: (AJ005671) cytochrome b6f complex subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 46
- Location of Alignment in SEQ ID NO 1910: from 1 to 46

Maximum Length Sequence:
related to:
Clone IDs:
31129
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1911
- Ceres seq_id 1571360
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1912
- Ceres seq_id 1571361
- Location of start within SEQ ID NO 1911: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 868
- gi No. 2804409
- Description: (AF043692) No definition line found [Caenorhabditis elegans]
- % Identity: 72.5
- Alignment Length: 41
- Location of Alignment in SEQ ID NO 1912: from 28 to 66

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1913
- Ceres seq_id 1571362
- Location of start within SEQ ID NO 1911: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1914
- Ceres seq_id 1571363
- Location of start within SEQ ID NO 1911: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 869
- gi No. 2804409
- Description: (AF043692) No definition line found [Caenorhabditis elegans]
- % Identity: 72.5
- Alignment Length: 41

- Location of Alignment in SEQ ID NO 1914: from 13 to 51

Maximum Length Sequence:
related to:
Clone IDs:
31538
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1915
- Ceres seq_id 1571364
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1916
- Ceres seq_id 1571365
- Location of start within SEQ ID NO 1915: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L15
- Location within SEQ ID NO 1916: from 2 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 870
- gi No. 3122673
- Description: 60S RIBOSOMAL PROTEIN L15
>gi|2245027|emb|CAB10447.1| (Z97341) ribosomal protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 204
- Location of Alignment in SEQ ID NO 1916: from 1 to 204

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1917
- Ceres seq_id 1571366
- Location of start within SEQ ID NO 1915: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L15
- Location within SEQ ID NO 1917: from 1 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 871
- gi No. 3122673
- Description: 60S RIBOSOMAL PROTEIN L15
>gi|2245027|emb|CAB10447.1| (Z97341) ribosomal protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 204
- Location of Alignment in SEQ ID NO 1917: from 1 to 186

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1918
- Ceres seq_id 1571367
- Location of start within SEQ ID NO 1915: at 446 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
31923
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 1919
- Ceres seq_id 1571372

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1920
- Ceres seq_id 1571373
- Location of start within SEQ ID NO 1919: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 872
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana]
- % Identity: 81.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 1920: from 1 to 141

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1921
- Ceres seq_id 1571374
- Location of start within SEQ ID NO 1919: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 873
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana]
- % Identity: 81.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 1921: from 1 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1922
- Ceres seq_id 1571375
- Location of start within SEQ ID NO 1919: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 874
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana|
- % Identity: 81.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 1922: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
3250

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1923
- Ceres seq_id 1571376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1924
- Ceres seq_id 1571377
- Location of start within SEQ ID NO 1923: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 1924: from 47 to 322 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1925
- Ceres seq_id 1571378
- Location of start within SEQ ID NO 1923: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 1925: from 5 to 280 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1926
- Ceres seq_id 1571379
- Location of start within SEQ ID NO 1923: at 326 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 1926: from 1 to 238 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3312
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1927
- Ceres seq_id 1571387

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1928
- Ceres seq_id 1571388
- Location of start within SEQ ID NO 1927: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Macrophage migration inhibitory factor (MIF)
- Location within SEQ ID NO 1928: from 28 to 141 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1929
- Ceres seq_id 1571389
- Location of start within SEQ ID NO 1927: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Macrophage migration inhibitory factor (MIF)
- Location within SEQ ID NO 1929: from 2 to 115 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1930
- Ceres seq_id 1571390
- Location of start within SEQ ID NO 1927: at 196 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Macrophage migration inhibitory factor (MIF)
- Location within SEQ ID NO 1930: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
33333
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1931
- Ceres seq_id 1571391
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1932
- Ceres seq_id 1571392
- Location of start within SEQ ID NO 1931: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 1932: from 14 to 61 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 875
- gi No. 3080388
- Description: (AL022603) myb-related protein [Arabidopsis thaliana] >gi|3402763|emb|CAA20209.1| (AL031187) myb-related protein M4 [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 191
- Location of Alignment in SEQ ID NO 1932: from 1 to 189

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1933
- Ceres seq_id 1571393
- Location of start within SEQ ID NO 1931: at 803 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 876
- gi No. 3080388
- Description: (AL022603) myb-related protein [Arabidopsis thaliana] >gi|3402763|emb|CAA20209.1| (AL031187) myb-related protein M4 [Arabidopsis thaliana]
- % Identity: 97
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 1933: from 1 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1934
- Ceres seq_id 1571394

- Location of start within SEQ ID NO 1931: at 842 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 877
- gi No. 3080388
- Description: (AL022603) myb-related protein [Arabidopsis thaliana] >gi|3402763|emb|CAA20209.1| (AL031187) myb-related protein M4 [Arabidopsis thaliana]
- % Identity: 97
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 1934: from 1 to 106

Maximum Length Sequence:
related to:
Clone IDs:
33382
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1935
- Ceres seq_id 1571395
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1936
- Ceres seq_id 1571396
- Location of start within SEQ ID NO 1935: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 878
- gi No. 1213594
- Description: (X93080) responsible for fatty acid elongation from C28 to C30 [Arabidopsis thaliana] >gi|1655786 (U40849) CER2 gene product [Arabidopsis thaliana] >gi|4220539|emb|CAA23012| (AL035356) CER2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 421
- Location of Alignment in SEQ ID NO 1936: from 31 to 451

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1937
- Ceres seq_id 1571397
- Location of start within SEQ ID NO 1935: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 879
- gi No. 1213594
- Description: (X93080) responsible for fatty acid elongation from C28 to C30 [Arabidopsis thaliana] >gi|1655786 (U40849) CER2 gene product [Arabidopsis thaliana] >gi|4220539|emb|CAA23012| (AL035356) CER2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 421
- Location of Alignment in SEQ ID NO 1937: from 1 to 421

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1938
- Ceres seq_id 1571398
- Location of start within SEQ ID NO 1935: at 181 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 880
- gi No. 1213594
- Description: (X93080) responsible for fatty acid elongation from C28 to C30 [Arabidopsis thaliana] >gi|1655786 (U40849) CER2 gene product [Arabidopsis thaliana] >gi|4220539|emb|CAA23012| (AL035356) CER2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 421
- Location of Alignment in SEQ ID NO 1938: from 1 to 391

Maximum Length Sequence:
related to:
Clone IDs:
34480

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1939
- Ceres seq_id 1571418

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1940
- Ceres seq_id 1571419
- Location of start within SEQ ID NO 1939: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 1940: from 89 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 881
- gi No. 5360591
- Description: (AB022328) nClpP3 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 1940: from 21 to 328

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1941
- Ceres seq_id 1571420
- Location of start within SEQ ID NO 1939: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 1941: from 68 to 256 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 882
- gi No. 5360591
- Description: (AB022328) nClpP3 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 1941: from 1 to 307

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1942
- Ceres seq_id 1571421
- Location of start within SEQ ID NO 1939: at 196 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 1942: from 24 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 883
- gi No. 5360591
- Description: (AB022328) nClpP3 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 308
- Location of Alignment in SEQ ID NO 1942: from 1 to 263

Maximum Length Sequence:
related to:
Clone IDs:
34599
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1943
- Ceres seq_id 1571422
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1944
- Ceres seq_id 1571423
- Location of start within SEQ ID NO 1943: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1944: from 57 to 268 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 884
- gi No. 3668069
- Description: (U28007) Pto kinase interactor 1 [Lycopersicon esculentum]
- % Identity: 73.2
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 1944: from 15 to 339

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1945
- Ceres seq_id 1571424
- Location of start within SEQ ID NO 1943: at 406 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 1945: from 1 to 133 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 885
- gi No. 3668069
- Description: (U28007) Pto kinase interactor 1 [Lycopersicon esculentum]
- % Identity: 73.2
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 1945: from 1 to 204

Maximum Length Sequence:
related to:

Clone IDs:
34671
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1946
- Ceres seq_id 1571425
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1947
- Ceres seq_id 1571426
- Location of start within SEQ ID NO 1946: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclic nucleotide-binding domain
- Location within SEQ ID NO 1947: from 24 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 886
- gi No. 3047124
- Description: (AF058919) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 1947: from 173 to 240

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1948
- Ceres seq_id 1571427
- Location of start within SEQ ID NO 1946: at 27 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclic nucleotide-binding domain
- Location within SEQ ID NO 1948: from 16 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 887
- gi No. 3047124
- Description: (AF058919) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 1948: from 165 to 232

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1949
- Ceres seq_id 1571428
- Location of start within SEQ ID NO 1946: at 345 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 888
- gi No. 3047124
- Description: (AF058919) No definition line found [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 1949: from 59 to 126

Maximum Length Sequence:
related to:

Clone IDs:
34936
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1950
- Ceres seq_id 1571432
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1951
- Ceres seq_id 1571433
- Location of start within SEQ ID NO 1950: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1951: from 72 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 889
- gi No. 3249084
- Description: (AC004473) Similar to red-1 (related to thioredoxin) gene gb|X92750 from Mus musculus. ESTs gb|AA712687 and gb|Z37223 come from this gene [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 443
- Location of Alignment in SEQ ID NO 1951: from 25 to 467

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1952
- Ceres seq_id 1571434
- Location of start within SEQ ID NO 1950: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1952: from 48 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 890
- gi No. 3249084
- Description: (AC004473) Similar to red-1 (related to thioredoxin) gene gb|X92750 from Mus musculus. ESTs gb|AA712687 and gb|Z37223 come from this gene [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 443
- Location of Alignment in SEQ ID NO 1952: from 1 to 443

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1953
- Ceres seq_id 1571435
- Location of start within SEQ ID NO 1950: at 367 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 1953: from 270 to 297 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 891
- gi No. 3249084
- Description: (AC004473) Similar to red-1 (related to thioredoxin) gene gb|X92750 from Mus musculus. ESTs gb|AA712687 and gb|Z37223 come from this gene [Arabidopsis thaliana]

- % Identity: 99.3
- Alignment Length: 443
- Location of Alignment in SEQ ID NO 1953: from 1 to 345

Maximum Length Sequence:
related to:
Clone IDs:
35228
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1954
- Ceres seq_id 1571440
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1955
- Ceres seq_id 1571441
- Location of start within SEQ ID NO 1954: at 208 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 1955: from 86 to 156 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 892
- gi No. 1781299
- Description: (Y09506) transformer-SR ribonucleoprotein [Nicotiana tabacum]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 1955: from 229 to 239

Maximum Length Sequence:
related to:
Clone IDs:
35385
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1956
- Ceres seq_id 1571445
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1957
- Ceres seq_id 1571446
- Location of start within SEQ ID NO 1956: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1957: from 86 to 482 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 893
- gi No. 4850392
- Description: (AC007357) Strong similarity to gb|X97864 cytochrome P450 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family. ESTs gb|N65665, gb|T14112, gb|T76255, gb|T20906 and gb|AI100027 come from ...
- % Identity: 97.6
- Alignment Length: 413
- Location of Alignment in SEQ ID NO 1957: from 86 to 498

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1958
- Ceres seq_id 1571447

- Location of start within SEQ ID NO 1956: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1958: from 53 to 449 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 894
- gi No. 4850392
- Description: (AC007357) Strong similarity to gb|X97864 cytochrome P450 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family. ESTs gb|N65665, gb|T14112, gb|T76255, gb|T20906 and gb|AI100027 come from ...
- % Identity: 97.6
- Alignment Length: 413
- Location of Alignment in SEQ ID NO 1958: from 53 to 465

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1959
- Ceres seq_id 1571448
- Location of start within SEQ ID NO 1956: at 393 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1959: from 1 to 352 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 895
- gi No. 4850392
- Description: (AC007357) Strong similarity to gb|X97864 cytochrome P450 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome P450 family. ESTs gb|N65665, gb|T14112, gb|T76255, gb|T20906 and gb|AI100027 come from ...
- % Identity: 97.6
- Alignment Length: 413
- Location of Alignment in SEQ ID NO 1959: from 1 to 368

Maximum Length Sequence:
related to:
Clone IDs:
35731
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1960
- Ceres seq_id 1571460
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1961
- Ceres seq_id 1571461
- Location of start within SEQ ID NO 1960: at 194 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1961: from 70 to 106 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1962
- Ceres seq_id 1571462

- Location of start within SEQ ID NO 1960: at 407 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1962: from 1 to 35 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1963
- Ceres seq_id 1571463
- Location of start within SEQ ID NO 1960: at 437 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
35786
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1964
- Ceres seq_id 1571464
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1965
- Ceres seq_id 1571465
- Location of start within SEQ ID NO 1964: at 144 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 1965: from 126 to 154 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 896
- gi No. 2827544
- Description: (AL021635) HSP associated protein like [Arabidopsis thaliana]
- % Identity: 95
- Alignment Length: 301
- Location of Alignment in SEQ ID NO 1965: from 38 to 338

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1966
- Ceres seq_id 1571466
- Location of start within SEQ ID NO 1964: at 342 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 1966: from 60 to 88 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 897
- gi No. 2827544
- Description: (AL021635) HSP associated protein like [Arabidopsis thaliana]
- % Identity: 95
- Alignment Length: 301

- Location of Alignment in SEQ ID NO 1966: from 1 to 272

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1967
- Ceres seq_id 1571467
- Location of start within SEQ ID NO 1964: at 465 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 1967: from 19 to 47 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 898
- gi No. 2827544
- Description: (AL021635) HSP associated protein like [Arabidopsis thaliana]
- % Identity: 95
- Alignment Length: 301
- Location of Alignment in SEQ ID NO 1967: from 1 to 231

Maximum Length Sequence:
related to:
Clone IDs:
35989
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1968
- Ceres seq_id 1571468
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1969
- Ceres seq_id 1571469
- Location of start within SEQ ID NO 1968: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 1969: from 27 to 108 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 899
- gi No. 2129769
- Description: xyloglucan endo-transglycosylase precursor - Arabidopsis thaliana >gi|944810|dbj|BAA09783| (D63508) endo-xyloglucan transferase [Arabidopsis thaliana] >gi|5730137|emb|CAB52471.1| (AL109796) xyloglucan endo-1,
- % Identity: 98.9
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 1969: from 20 to 108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1970
- Ceres seq_id 1571470
- Location of start within SEQ ID NO 1968: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 1970: from 8 to 89 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 900

- gi No. 2129769
- Description: xyloglucan endo-transglycosylase precursor - Arabidopsis thaliana >gi|944810|dbj|BAA09783| (D63508) endo-xyloglucan transferase [Arabidopsis thaliana] >gi|5730137|emb|CAB52471.1| (AL109796) xyloglucan endo-1,
- % Identity: 98.9
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 1970: from 1 to 89

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1971
- Ceres seq_id 1571471
- Location of start within SEQ ID NO 1968: at 470 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 16
- Location within SEQ ID NO 1971: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 901
- gi No. 2129769
- Description: xyloglucan endo-transglycosylase precursor - Arabidopsis thaliana >gi|944810|dbj|BAA09783| (D63508) endo-xyloglucan transferase [Arabidopsis thaliana] >gi|5730137|emb|CAB52471.1| (AL109796) xyloglucan endo-1,
- % Identity: 100
- Alignment Length: 181
- Location of Alignment in SEQ ID NO 1971: from 1 to 163

Maximum Length Sequence:
related to:
Clone IDs:
36023
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1972
- Ceres seq_id 1571472

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1973
- Ceres seq_id 1571473
- Location of start within SEQ ID NO 1972: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 902
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 97
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 1973: from 147 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1974
- Ceres seq_id 1571474
- Location of start within SEQ ID NO 1972: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 903
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 97
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 1974: from 132 to 163

Maximum Length Sequence:
related to:
Clone IDs:
36216
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1975
- Ceres seq_id 1571479
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1976
- Ceres seq_id 1571480
- Location of start within SEQ ID NO 1975: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1976: from 54 to 307 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 904
- gi No. 2462824
- Description: (AF000657) similar to Jun activation domain binding protein [Arabidopsis thaliana] >gi|2791885 (AF042334) JAB1 [Arabidopsis thaliana]
- % Identity: 97.1
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 1976: from 1 to 307

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1977
- Ceres seq_id 1571481
- Location of start within SEQ ID NO 1975: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1977: from 1 to 239 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 905
- gi No. 2462824
- Description: (AF000657) similar to Jun activation domain binding protein [Arabidopsis thaliana] >gi|2791885 (AF042334) JAB1 [Arabidopsis thaliana]
- % Identity: 97.1
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 1977: from 1 to 239

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1978
- Ceres seq_id 1571482
- Location of start within SEQ ID NO 1975: at 404 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1978: from 1 to 226 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 906
- gi No. 2462824
- Description: (AF000657) similar to Jun activation domain binding protein [Arabidopsis thaliana] >gi|2791885 (AF042334) JAB1 [Arabidopsis thaliana]
- % Identity: 97.1
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 1978: from 1 to 226

Maximum Length Sequence:
related to:
Clone IDs:
36325
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1979
- Ceres seq_id 1571483
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1980
- Ceres seq_id 1571484
- Location of start within SEQ ID NO 1979: at 171 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 907
- gi No. 998671
- Description: acetyl esterase {N-terminal} [Vigna radiata=mung beans, Wilczek, hypocotyls, Peptide Partial, 31 aa, segment 1 of 2]
- % Identity: 75.9
- Alignment Length: 29
- Location of Alignment in SEQ ID NO 1980: from 25 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1981
- Ceres seq_id 1571485
- Location of start within SEQ ID NO 1979: at 222 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 908
- gi No. 998671
- Description: acetyl esterase {N-terminal} [Vigna radiata=mung beans, Wilczek, hypocotyls, Peptide Partial, 31 aa, segment 1 of 2]
- % Identity: 75.9
- Alignment Length: 29
- Location of Alignment in SEQ ID NO 1981: from 8 to 36

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1982
- Ceres seq_id 1571486

- Location of start within SEQ ID NO 1979: at 411 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 909
- gi No. 998672
- Description: acetyl esterase {N-terminal} [Vigna radiata=mung beans, Wilczek, hypocotyls, Peptide Partial, 23 aa, segment 2 of 2]
- % Identity: 85.7
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1982: from 43 to 63

Maximum Length Sequence:
related to:
Clone IDs:
36334
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1983
- Ceres seq_id 1571487
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1984
- Ceres seq_id 1571488
- Location of start within SEQ ID NO 1983: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1984: from 52 to 487 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 910
- gi No. 1076315
- Description: cytochrome P450 - Arabidopsis thaliana
>gi|853719|emb|CAA60793| (X87367) CYP90 protein [Arabidopsis thaliana]
>gi|871988|emb|CAA60794| (X87368) CYP90 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 472
- Location of Alignment in SEQ ID NO 1984: from 21 to 492

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1985
- Ceres seq_id 1571489
- Location of start within SEQ ID NO 1983: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1985: from 32 to 467 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 911
- gi No. 1076315
- Description: cytochrome P450 - Arabidopsis thaliana
>gi|853719|emb|CAA60793| (X87367) CYP90 protein [Arabidopsis thaliana]
>gi|871988|emb|CAA60794| (X87368) CYP90 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 472
- Location of Alignment in SEQ ID NO 1985: from 1 to 472

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 1986
- Ceres seq_id 1571490
- Location of start within SEQ ID NO 1983: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 1986: from 4 to 439 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 912
- gi No. 1076315
- Description: cytochrome P450 - Arabidopsis thaliana >gi|853719|emb|CAA60793| (X87367) CYP90 protein [Arabidopsis thaliana] >gi|871988|emb|CAA60794| (X87368) CYP90 protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 472
- Location of Alignment in SEQ ID NO 1986: from 1 to 444

Maximum Length Sequence:
related to:
Clone IDs:
36577
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1987
- Ceres seq_id 1571499

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1988
- Ceres seq_id 1571500
- Location of start within SEQ ID NO 1987: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1988: from 140 to 192 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 913
- gi No. 3150415
- Description: (AC004165) sec13-related protein [Arabidopsis thaliana] >gi|3420046 (AC004680) sec13-related protein [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 302
- Location of Alignment in SEQ ID NO 1988: from 1 to 302

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1989
- Ceres seq_id 1571501
- Location of start within SEQ ID NO 1987: at 208 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 1989: from 122 to 174 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 914
- gi No. 3150415
- Description: (AC004165) sec13-related protein [Arabidopsis thaliana] >gi|3420046 (AC004680) sec13-related protein [Arabidopsis thaliana]
- % Identity: 99.7

- Alignment Length: 302
- Location of Alignment in SEQ ID NO 1989: from 1 to 284

Maximum Length Sequence:
related to:
Clone IDs:
36616
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1990
- Ceres seq_id 1571502
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1991
- Ceres seq_id 1571503
- Location of start within SEQ ID NO 1990: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1991: from 8 to 231 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1992
- Ceres seq_id 1571504
- Location of start within SEQ ID NO 1990: at 8 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1992: from 6 to 229 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1993
- Ceres seq_id 1571505
- Location of start within SEQ ID NO 1990: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 1993: from 1 to 212 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
36882
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1994
- Ceres seq_id 1571522
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1995
- Ceres seq_id 1571523
- Location of start within SEQ ID NO 1994: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 14

- Location within SEQ ID NO 1995: from 117 to 535 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 915
- gi No. 404372
- Description: beta-amylase {N-terminal} [Glycine max=soybeans, Merrill, cv Provar, seeds, Peptide Partial, 27 aa]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1995: from 273 to 293

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1996
- Ceres seq_id 1571524
- Location of start within SEQ ID NO 1994: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 14
- Location within SEQ ID NO 1996: from 89 to 507 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 916
- gi No. 404372
- Description: beta-amylase {N-terminal} [Glycine max=soybeans, Merrill, cv Provar, seeds, Peptide Partial, 27 aa]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1996: from 245 to 265

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1997
- Ceres seq_id 1571525
- Location of start within SEQ ID NO 1994: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 14
- Location within SEQ ID NO 1997: from 60 to 478 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 917
- gi No. 404372
- Description: beta-amylase {N-terminal} [Glycine max=soybeans, Merrill, cv Provar, seeds, Peptide Partial, 27 aa]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 1997: from 216 to 236

Maximum Length Sequence:
related to:
Clone IDs:
37019
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 1998
- Ceres seq_id 1571529
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 1999
- Ceres seq_id 1571530
- Location of start within SEQ ID NO 1998: at 228 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 918
- gi No. 4966357
- Description: (AC006341) >F3O9.16 [Arabidopsis thaliana]
- % Identity: 75.2
- Alignment Length: 274
- Location of Alignment in SEQ ID NO 1999: from 76 to 349

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2000
- Ceres seq_id 1571531
- Location of start within SEQ ID NO 1998: at 579 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 919
- gi No. 4966357
- Description: (AC006341) >F3O9.16 [Arabidopsis thaliana]
- % Identity: 75.2
- Alignment Length: 274
- Location of Alignment in SEQ ID NO 2000: from 1 to 232

Maximum Length Sequence:
related to:
Clone IDs:
37206
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2001
- Ceres seq_id 1571543
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2002
- Ceres seq_id 1571544
- Location of start within SEQ ID NO 2001: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 2002: from 82 to 347 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2003
- Ceres seq_id 1571545
- Location of start within SEQ ID NO 2001: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 2003: from 73 to 338 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2004
- Ceres seq_id 1571546
- Location of start within SEQ ID NO 2001: at 224 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cyclin
- Location within SEQ ID NO 2004: from 61 to 326 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3721
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2005
- Ceres seq_id 1571547
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2006
- Ceres seq_id 1571548
- Location of start within SEQ ID NO 2005: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2006: from 50 to 269 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2007
- Ceres seq_id 1571549
- Location of start within SEQ ID NO 2005: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2007: from 1 to 215 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2008
- Ceres seq_id 1571550
- Location of start within SEQ ID NO 2005: at 410 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2008: from 1 to 154 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37288
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2009
- Ceres seq_id 1571559
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2010
- Ceres seq_id 1571560
- Location of start within SEQ ID NO 2009: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 2010: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 920
- gi No. 1362199
- Description: homeotic protein ZAG2 - maize (fragment)
- % Identity: 74.6
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 2010: from 1 to 59

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2011
- Ceres seq_id 1571561
- Location of start within SEQ ID NO 2009: at 394 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2012
- Ceres seq_id 1571562
- Location of start within SEQ ID NO 2009: at 520 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37341
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2013
- Ceres seq_id 1571567
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2014
- Ceres seq_id 1571568
- Location of start within SEQ ID NO 2013: at 55 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- haloacid dehalogenase-like hydrolase
- Location within SEQ ID NO 2014: from 63 to 135 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2015
- Ceres seq_id 1571569
- Location of start within SEQ ID NO 2013: at 427 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- haloacid dehalogenase-like hydrolase
- Location within SEQ ID NO 2015: from 37 to 153 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37538
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2016
- Ceres seq_id 1571570
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2017
- Ceres seq_id 1571571
- Location of start within SEQ ID NO 2016: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2017: from 33 to 506 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 921
- gi No. 2252844
- Description: (AF013293) belongs to the cytochrome p450 family [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 525
- Location of Alignment in SEQ ID NO 2017: from 6 to 525

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2018
- Ceres seq_id 1571572
- Location of start within SEQ ID NO 2016: at 16 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2018: from 28 to 501 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 922
- gi No. 2252844
- Description: (AF013293) belongs to the cytochrome p450 family [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 525
- Location of Alignment in SEQ ID NO 2018: from 1 to 520

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2019
- Ceres seq_id 1571573
- Location of start within SEQ ID NO 2016: at 157 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2019: from 1 to 454 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 923
- gi No. 2252844

- Description: (AF013293) belongs to the cytochrome p450 family [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 525
- Location of Alignment in SEQ ID NO 2019: from 1 to 473

Maximum Length Sequence:
related to:
Clone IDs:
37663
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2020
- Ceres seq_id 1571582
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2021
- Ceres seq_id 1571583
- Location of start within SEQ ID NO 2020: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 924
- gi No. 2262110
- Description: (AC002343) zinc finger protein isolog [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 2021: from 1 to 342

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2022
- Ceres seq_id 1571584
- Location of start within SEQ ID NO 2020: at 575 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 925
- gi No. 2262110
- Description: (AC002343) zinc finger protein isolog [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 2022: from 1 to 204

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2023
- Ceres seq_id 1571585
- Location of start within SEQ ID NO 2020: at 581 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 926
- gi No. 2262110
- Description: (AC002343) zinc finger protein isolog [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 342
- Location of Alignment in SEQ ID NO 2023: from 1 to 202

Maximum Length Sequence:
related to:
Clone IDs:
37699
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2024
- Ceres seq_id 1571586
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2025
- Ceres seq_id 1571587
- Location of start within SEQ ID NO 2024: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2025: from 31 to 433 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 927
- gi No. 3142294
- Description: (AC002411) Strong similarity to initiation factor eIF-2, gb|U37354 from S. pombe. ESTs gb|T41979, gb|N37284 and gb|N37529 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 2025: from 1 to 465

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2026
- Ceres seq_id 1571588
- Location of start within SEQ ID NO 2024: at 389 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2026: from 1 to 342 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 928
- gi No. 3142294
- Description: (AC002411) Strong similarity to initiation factor eIF-2, gb|U37354 from S. pombe. ESTs gb|T41979, gb|N37284 and gb|N37529 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 2026: from 1 to 374

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2027
- Ceres seq_id 1571589
- Location of start within SEQ ID NO 2024: at 464 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2027: from 1 to 317 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 929
- gi No. 3142294

- Description: (AC002411) Strong similarity to initiation factor eIF-2, gb|U37354 from S. pombe. ESTs gb|T41979, gb|N37284 and gb|N37529 come from this gene. [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 465
- Location of Alignment in SEQ ID NO 2027: from 1 to 349

Maximum Length Sequence:
related to:
Clone IDs:
37907
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2028
- Ceres seq_id 1571612

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2029
- Ceres seq_id 1571613
- Location of start within SEQ ID NO 2028: at 223 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 2029: from 84 to 324 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 930
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2029: from 1 to 364

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2030
- Ceres seq_id 1571614
- Location of start within SEQ ID NO 2028: at 364 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 2030: from 37 to 277 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 931
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2030: from 1 to 317

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2031
- Ceres seq_id 1571615
- Location of start within SEQ ID NO 2028: at 559 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Dihydrodipicolinate synthetase family
- Location within SEQ ID NO 2031: from 1 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 932
- gi No. 1076317
- Description: dihydrodipicolinate synthase (EC 4.2.1.52) precursor - Arabidopsis thaliana >gi|5139329|emb|CAB45642.1| (X72971) dihydrodipicolinate synthase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2031: from 1 to 252

Maximum Length Sequence:
related to:
Clone IDs:
37952
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2032
- Ceres seq_id 1571616
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2033
- Ceres seq_id 1571617
- Location of start within SEQ ID NO 2032: at 572 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectate lyase
- Location within SEQ ID NO 2033: from 1 to 267 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 933
- gi No. 2494113
- Description: (AC002376) Strong similarity to Musa pectate lyase (gb|X92943). ESTs gb|AA042458, gb|ATTS4502, gb|N38552 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 290
- Location of Alignment in SEQ ID NO 2033: from 1 to 276

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2034
- Ceres seq_id 1571618
- Location of start within SEQ ID NO 2032: at 740 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectate lyase
- Location within SEQ ID NO 2034: from 1 to 211 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 934
- gi No. 2494113
- Description: (AC002376) Strong similarity to Musa pectate lyase (gb|X92943). ESTs gb|AA042458, gb|ATTS4502, gb|N38552 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 290
- Location of Alignment in SEQ ID NO 2034: from 1 to 220

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2035
- Ceres seq_id 1571619
- Location of start within SEQ ID NO 2032: at 779 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pectate lyase
- Location within SEQ ID NO 2035: from 1 to 198 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 935
- gi No. 2494113
- Description: (AC002376) Strong similarity to Musa pectate lyase (gb|X92943). ESTs gb|AA042458, gb|ATTS4502, gb|N38552 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 290
- Location of Alignment in SEQ ID NO 2035: from 1 to 207

Maximum Length Sequence:
related to:
Clone IDs:
38185
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2036
- Ceres seq_id 1571628

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2037
- Ceres seq_id 1571629
- Location of start within SEQ ID NO 2036: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2037: from 41 to 182 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 936
- gi No. 3421104
- Description: (AF043531) 20S proteasome beta subunit PBB2 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 274
- Location of Alignment in SEQ ID NO 2037: from 1 to 274

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2038
- Ceres seq_id 1571630
- Location of start within SEQ ID NO 2036: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2038: from 18 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 937
- gi No. 3421104
- Description: (AF043531) 20S proteasome beta subunit PBB2 [Arabidopsis thaliana]
- % Identity: 99.6

- Alignment Length: 274
- Location of Alignment in SEQ ID NO 2038: from 1 to 251

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2039
- Ceres seq_id 1571631
- Location of start within SEQ ID NO 2036: at 375 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2039: from 1 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 938
- gi No. 3421104
- Description: (AF043531) 20S proteasome beta subunit PBB2 [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 274
- Location of Alignment in SEQ ID NO 2039: from 1 to 199

Maximum Length Sequence:
related to:
Clone IDs:
38601
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2040
- Ceres seq_id 1571642

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2041
- Ceres seq_id 1571643
- Location of start within SEQ ID NO 2040: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 939
- gi No. 5123545
- Description: (AL079344) arginine methyltransferase (pam1) [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 2041: from 1 to 390

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2042
- Ceres seq_id 1571644
- Location of start within SEQ ID NO 2040: at 262 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 940
- gi No. 5123545
- Description: (AL079344) arginine methyltransferase (pam1) [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 2042: from 1 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2043
- Ceres seq_id 1571645
- Location of start within SEQ ID NO 2040: at 367 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 941
- gi No. 5123545
- Description: (AL079344) arginine methyltransferase (pam1) [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 390
- Location of Alignment in SEQ ID NO 2043: from 1 to 306

Maximum Length Sequence:
related to:
Clone IDs:
38945

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2044
- Ceres seq_id 1571669

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2045
- Ceres seq_id 1571670
- Location of start within SEQ ID NO 2044: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2045: from 26 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 942
- gi No. 4033467
- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP31 >gi|1707366|emb|CAA67798| (X99435) splicing factor [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2045: from 23 to 211

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2046
- Ceres seq_id 1571671
- Location of start within SEQ ID NO 2044: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2046: from 4 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 943
- gi No. 4033467
- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP31 >gi|1707366|emb|CAA67798| (X99435) splicing factor [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2046: from 1 to 189

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2047
- Ceres seq_id 1571672
- Location of start within SEQ ID NO 2044: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2047: from 63 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 944
- gi No. 4033467
- Description: ARGININE/SERINE-RICH SPLICING FACTOR RSP31 >gi|1707366|emb|CAA67798| (X99435) splicing factor [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2047: from 1 to 157

Maximum Length Sequence:
related to:
Clone IDs:
39167

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2048
- Ceres seq_id 1571677

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2049
- Ceres seq_id 1571678
- Location of start within SEQ ID NO 2048: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- haloacid dehalogenase-like hydrolase
- Location within SEQ ID NO 2049: from 10 to 219 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 945
- gi No. 2660676
- Description: (AC002342) Dreg-2 like protein [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 255
- Location of Alignment in SEQ ID NO 2049: from 1 to 255

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2050
- Ceres seq_id 1571679
- Location of start within SEQ ID NO 2048: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- haloacid dehalogenase-like hydrolase
- Location within SEQ ID NO 2050: from 1 to 187 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 946
- gi No. 2660676
- Description: (AC002342) Dreg-2 like protein [Arabidopsis thaliana]
- % Identity: 99.6

- Alignment Length: 255
- Location of Alignment in SEQ ID NO 2050: from 1 to 223

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2051
- Ceres seq_id 1571680
- Location of start within SEQ ID NO 2048: at 289 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- haloacid dehalogenase-like hydrolase
- Location within SEQ ID NO 2051: from 1 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 947
- gi No. 2660676
- Description: (AC002342) Dreg-2 like protein [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 255
- Location of Alignment in SEQ ID NO 2051: from 1 to 197

Maximum Length Sequence:
related to:
Clone IDs:
39745
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2052
- Ceres seq_id 1571687
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2053
- Ceres seq_id 1571688
- Location of start within SEQ ID NO 2052: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 2053: from 75 to 451 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2054
- Ceres seq_id 1571689
- Location of start within SEQ ID NO 2052: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 2054: from 50 to 426 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2055
- Ceres seq_id 1571690
- Location of start within SEQ ID NO 2052: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter

- Location within SEQ ID NO 2055: from 20 to 396 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39762
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2056
- Ceres seq_id 1571691
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2057
- Ceres seq_id 1571692
- Location of start within SEQ ID NO 2056: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2057: from 42 to 188 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2058
- Ceres seq_id 1571693
- Location of start within SEQ ID NO 2056: at 45 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2058: from 28 to 174 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2059
- Ceres seq_id 1571694
- Location of start within SEQ ID NO 2056: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2059: from 13 to 159 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39872
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2060
- Ceres seq_id 1571703
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2061
- Ceres seq_id 1571704
- Location of start within SEQ ID NO 2060: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 948
- gi No. 2245123
- Description: (Z97343) sperm protein homolog [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2061: from 10 to 172

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2062
- Ceres seq_id 1571705
- Location of start within SEQ ID NO 2060: at 241 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 949
- gi No. 2245123
- Description: (Z97343) sperm protein homolog [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2062: from 6 to 168

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2063
- Ceres seq_id 1571706
- Location of start within SEQ ID NO 2060: at 403 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 950
- gi No. 2245123
- Description: (Z97343) sperm protein homolog [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2063: from 1 to 114

Maximum Length Sequence:
related to:
Clone IDs:
39882
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2064
- Ceres seq_id 1571707
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2065
- Ceres seq_id 1571708
- Location of start within SEQ ID NO 2064: at 379 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ion transport protein
- Location within SEQ ID NO 2065: from 62 to 132 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 951
- gi No. 2181186

- Description: (X97323) outward rectifying potassium channel KCO1 [Arabidopsis thaliana] >gi|2230761|emb|CAA69158| (Y07825) kco1 [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2065: from 1 to 253

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2066
- Ceres seq_id 1571709
- Location of start within SEQ ID NO 2064: at 469 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ion transport protein
- Location within SEQ ID NO 2066: from 32 to 102 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 952
- gi No. 2181186
- Description: (X97323) outward rectifying potassium channel KCO1 [Arabidopsis thaliana] >gi|2230761|emb|CAA69158| (Y07825) kco1 [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2066: from 1 to 223

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2067
- Ceres seq_id 1571710
- Location of start within SEQ ID NO 2064: at 526 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ion transport protein
- Location within SEQ ID NO 2067: from 13 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 953
- gi No. 2181186
- Description: (X97323) outward rectifying potassium channel KCO1 [Arabidopsis thaliana] >gi|2230761|emb|CAA69158| (Y07825) kco1 [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 364
- Location of Alignment in SEQ ID NO 2067: from 1 to 204

Maximum Length Sequence:
related to:
Clone IDs:
40423
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2068
- Ceres seq_id 1571734

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2069
- Ceres seq_id 1571735
- Location of start within SEQ ID NO 2068: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 954
- gi No. 4454026
- Description: (AL035394) phosphatase like protein [Arabidopsis thaliana]
- % Identity: 77
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2069: from 1 to 357

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2070
- Ceres seq_id 1571736
- Location of start within SEQ ID NO 2068: at 297 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 955
- gi No. 4454026
- Description: (AL035394) phosphatase like protein [Arabidopsis thaliana]
- % Identity: 77
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2070: from 1 to 277

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2071
- Ceres seq_id 1571737
- Location of start within SEQ ID NO 2068: at 381 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 956
- gi No. 4454026
- Description: (AL035394) phosphatase like protein [Arabidopsis thaliana]
- % Identity: 77
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2071: from 1 to 249

Maximum Length Sequence:
related to:
Clone IDs:
40576
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2072
- Ceres seq_id 1571750

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2073
- Ceres seq_id 1571751
- Location of start within SEQ ID NO 2072: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 2073: from 115 to 137 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 957
- gi No. 1418325

- Description: (X98674) zinc finger protein [Arabidopsis thaliana] >gi|1418337|emb|CAA67231| (X98673) zinc finger protein [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 2073: from 34 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2074
- Ceres seq_id 1571752
- Location of start within SEQ ID NO 2072: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 2074: from 82 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 958
- gi No. 1418325
- Description: (X98674) zinc finger protein [Arabidopsis thaliana] >gi|1418337|emb|CAA67231| (X98673) zinc finger protein [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 2074: from 1 to 109

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2075
- Ceres seq_id 1571753
- Location of start within SEQ ID NO 2072: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 2075: from 63 to 85 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 959
- gi No. 1418325
- Description: (X98674) zinc finger protein [Arabidopsis thaliana] >gi|1418337|emb|CAA67231| (X98673) zinc finger protein [Arabidopsis thaliana]
- % Identity: 99.1
- Alignment Length: 109
- Location of Alignment in SEQ ID NO 2075: from 1 to 90

Maximum Length Sequence:
related to:
Clone IDs:
40913
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2076
- Ceres seq_id 1571770

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2077
- Ceres seq_id 1571771
- Location of start within SEQ ID NO 2076: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 960
- gi No. 5302783

- Description: (Z97340) isomerase like protein [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2077: from 1 to 223

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2078
- Ceres seq_id 1571772
- Location of start within SEQ ID NO 2076: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 961
- gi No. 5302783
- Description: (Z97340) isomerase like protein [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2078: from 1 to 219

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2079
- Ceres seq_id 1571773
- Location of start within SEQ ID NO 2076: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 962
- gi No. 5302783
- Description: (Z97340) isomerase like protein [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2079: from 1 to 215

Maximum Length Sequence:
related to:
Clone IDs:
41397

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2080
- Ceres seq_id 1571788

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2081
- Ceres seq_id 1571789
- Location of start within SEQ ID NO 2080: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2081: from 45 to 330 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 963
- gi No. 2245128
- Description: (Z97344) peroxidase like protein [Arabidopsis thaliana]
- % Identity: 79.8

- Alignment Length: 321
- Location of Alignment in SEQ ID NO 2081: from 14 to 334

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2082
- Ceres seq_id 1571790
- Location of start within SEQ ID NO 2080: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2082: from 42 to 327 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 964
- gi No. 2245128
- Description: (Z97344) peroxidase like protein [Arabidopsis thaliana]
- % Identity: 79.8
- Alignment Length: 321
- Location of Alignment in SEQ ID NO 2082: from 11 to 331

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2083
- Ceres seq_id 1571791
- Location of start within SEQ ID NO 2080: at 527 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2083: from 1 to 188 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 965
- gi No. 2245128
- Description: (Z97344) peroxidase like protein [Arabidopsis thaliana]
- % Identity: 79.8
- Alignment Length: 321
- Location of Alignment in SEQ ID NO 2083: from 1 to 192

Maximum Length Sequence:
related to:
Clone IDs:
41446
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2084
- Ceres seq_id 1571792

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2085
- Ceres seq_id 1571793
- Location of start within SEQ ID NO 2084: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2085: from 44 to 330 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 966
- gi No. 1076635

- Description: protoporphyrinogen oxidase (EC 1.3.3.4) - common tobacco (fragment)
- % Identity: 73.3
- Alignment Length: 15
- Location of Alignment in SEQ ID NO 2085: from 64 to 78

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2086
- Ceres seq_id 1571794
- Location of start within SEQ ID NO 2084: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2086: from 35 to 321 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 967
- gi No. 1076635
- Description: protoporphyrinogen oxidase (EC 1.3.3.4) - common tobacco (fragment)
- % Identity: 73.3
- Alignment Length: 15
- Location of Alignment in SEQ ID NO 2086: from 55 to 69

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2087
- Ceres seq_id 1571795
- Location of start within SEQ ID NO 2084: at 206 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2087: from 1 to 262 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 968
- gi No. 1076635
- Description: protoporphyrinogen oxidase (EC 1.3.3.4) - common tobacco (fragment)
- % Identity: 83.3
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2087: from 122 to 133

Maximum Length Sequence:
related to:
Clone IDs:
41558
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2088
- Ceres seq_id 1571796

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2089
- Ceres seq_id 1571797
- Location of start within SEQ ID NO 2088: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Translation initiation factor IF-3
- Location within SEQ ID NO 2089: from 118 to 234 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 969
- gi No. 3924841
- Description: (Z66565) cDNA EST EMBL:D70617 comes from this gene; cDNA EST EMBL:C10400 comes from this gene; cDNA EST EMBL:C10605 comes from this gene; cDNA EST EMBL:C12377 comes from this gene; cDNA EST EMBL:C13473 comes from this gene; cDN...
- % Identity: 79.2
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 2089: from 284 to 307

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2090
- Ceres seq_id 1571798
- Location of start within SEQ ID NO 2088: at 391 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Translation initiation factor IF-3
- Location within SEQ ID NO 2090: from 10 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 970
- gi No. 3924841
- Description: (Z66565) cDNA EST EMBL:D70617 comes from this gene; cDNA EST EMBL:C10400 comes from this gene; cDNA EST EMBL:C10605 comes from this gene; cDNA EST EMBL:C12377 comes from this gene; cDNA EST EMBL:C13473 comes from this gene; cDN...
- % Identity: 79.2
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 2090: from 176 to 199

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2091
- Ceres seq_id 1571799
- Location of start within SEQ ID NO 2088: at 427 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Translation initiation factor IF-3
- Location within SEQ ID NO 2091: from 1 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 971
- gi No. 3924841
- Description: (Z66565) cDNA EST EMBL:D70617 comes from this gene; cDNA EST EMBL:C10400 comes from this gene; cDNA EST EMBL:C10605 comes from this gene; cDNA EST EMBL:C12377 comes from this gene; cDNA EST EMBL:C13473 comes from this gene; cDN...
- % Identity: 79.2
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 2091: from 164 to 187

Maximum Length Sequence:
related to:
Clone IDs:
41573
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2092
- Ceres seq_id 1571800
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2093
- Ceres seq_id 1571801
- Location of start within SEQ ID NO 2092: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldose 1-epimerase
- Location within SEQ ID NO 2093: from 20 to 105 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2094
- Ceres seq_id 1571802
- Location of start within SEQ ID NO 2092: at 278 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldose 1-epimerase
- Location within SEQ ID NO 2094: from 8 to 256 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2095
- Ceres seq_id 1571803
- Location of start within SEQ ID NO 2092: at 506 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aldose 1-epimerase
- Location within SEQ ID NO 2095: from 1 to 180 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
426
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2096
- Ceres seq_id 1571812

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2097
- Ceres seq_id 1571813
- Location of start within SEQ ID NO 2096: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2097: from 24 to 64 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2098
- Ceres seq_id 1571814
- Location of start within SEQ ID NO 2096: at 9 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2098: from 22 to 62 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2099
- Ceres seq_id 1571815
- Location of start within SEQ ID NO 2096: at 25 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
42932
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2100
- Ceres seq_id 1571820
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2101
- Ceres seq_id 1571821
- Location of start within SEQ ID NO 2100: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin family
- Location within SEQ ID NO 2101: from 2 to 221 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 972
- gi No. 2129608
- Description: GTP-binding protein, 68K - Arabidopsis thaliana >gi|807577 (L38614) GTP-binding protein [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 2101: from 1 to 429

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2102
- Ceres seq_id 1571822
- Location of start within SEQ ID NO 2100: at 713 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin family
- Location within SEQ ID NO 2102: from 1 to 57 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 973
- gi No. 2129608
- Description: GTP-binding protein, 68K - Arabidopsis thaliana >gi|807577 (L38614) GTP-binding protein [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 2102: from 1 to 265

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2103

- Ceres seq_id 1571823
- Location of start within SEQ ID NO 2100: at 866 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynamin central region
- Location within SEQ ID NO 2103: from 7 to 214 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 974
- gi No. 2129608
- Description: GTP-binding protein, 68K - Arabidopsis thaliana >gi|807577 (L38614) GTP-binding protein [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 429
- Location of Alignment in SEQ ID NO 2103: from 1 to 214

Maximum Length Sequence:
related to:
Clone IDs:
42992
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2104
- Ceres seq_id 1571824
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2105
- Ceres seq_id 1571825
- Location of start within SEQ ID NO 2104: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 975
- gi No. 5441887
- Description: (AP000367) Similar to Arabidopsis thaliana BAC genomic sequence. (AC002332) [Oryza sativa]
- % Identity: 71.2
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2105: from 26 to 188

Maximum Length Sequence:
related to:
Clone IDs:
43073
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2106
- Ceres seq_id 1571826
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2107
- Ceres seq_id 1571827
- Location of start within SEQ ID NO 2106: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxyl transferase domain
- Location within SEQ ID NO 2107: from 60 to 529 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 976
- gi No. 600178

- Description: (L27074) acetyl-CoA carboxylase [Arabidopsis thaliana]>gi|1090217|prf||2018327A Ac-CoA carboxylase [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 520
- Location of Alignment in SEQ ID NO 2107: from 11 to 529

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2108
- Ceres seq_id 1571828
- Location of start within SEQ ID NO 2106: at 242 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxyl transferase domain
- Location within SEQ ID NO 2108: from 1 to 465 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 977
- gi No. 600178
- Description: (L27074) acetyl-CoA carboxylase [Arabidopsis thaliana]>gi|1090217|prf||2018327A Ac-CoA carboxylase [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 520
- Location of Alignment in SEQ ID NO 2108: from 1 to 465

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2109
- Ceres seq_id 1571829
- Location of start within SEQ ID NO 2106: at 263 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Carboxyl transferase domain
- Location within SEQ ID NO 2109: from 1 to 458 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 978
- gi No. 600178
- Description: (L27074) acetyl-CoA carboxylase [Arabidopsis thaliana]>gi|1090217|prf||2018327A Ac-CoA carboxylase [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 520
- Location of Alignment in SEQ ID NO 2109: from 1 to 458

Maximum Length Sequence:
related to:
Clone IDs:
4325
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2110
- Ceres seq_id 1571830

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2111
- Ceres seq_id 1571831
- Location of start within SEQ ID NO 2110: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 2111: from 74 to 102 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2112
- Ceres seq_id 1571832
- Location of start within SEQ ID NO 2110: at 13 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 2112: from 70 to 98 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
4372
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2113
- Ceres seq_id 1571837
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2114
- Ceres seq_id 1571838
- Location of start within SEQ ID NO 2113: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2114: from 75 to 402 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 979
- gi No. 99696
- Description: glutamate--ammonia ligase (EC 6.3.1.2) precursor, chloroplast (clone lambdaAtgsl1) - Arabidopsis thaliana >gi|240070|bbs|69728 (S69727) light-regulated glutamine synthetase isoenzyme [Arabidopsis thaliana, Peptide, 430 aa] [Arabidopsis
- % Identity: 100
- Alignment Length: 430
- Location of Alignment in SEQ ID NO 2114: from 1 to 430

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2115
- Ceres seq_id 1571839
- Location of start within SEQ ID NO 2113: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2115: from 63 to 390 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 980
- gi No. 99696
- Description: glutamate--ammonia ligase (EC 6.3.1.2) precursor, chloroplast (clone lambdaAtgsl1) - Arabidopsis thaliana >gi|240070|bbs|69728 (S69727) light-regulated glutamine synthetase isoenzyme [Arabidopsis thaliana, Peptide, 430 aa] [Arabidopsis
- % Identity: 100
- Alignment Length: 430

- Location of Alignment in SEQ ID NO 2115: from 1 to 418

Maximum Length Sequence:
related to:
Clone IDs:
4581
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2116
- Ceres seq_id 1571840
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2117
- Ceres seq_id 1571841
- Location of start within SEQ ID NO 2116: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2117: from 119 to 160 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2118
- Ceres seq_id 1571842
- Location of start within SEQ ID NO 2116: at 218 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2118: from 94 to 135 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2119
- Ceres seq_id 1571843
- Location of start within SEQ ID NO 2116: at 233 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2119: from 89 to 130 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
4611
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2120
- Ceres seq_id 1571844
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2121
- Ceres seq_id 1571845
- Location of start within SEQ ID NO 2120: at 157 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GDP dissociation inhibitor
- Location within SEQ ID NO 2121: from 1 to 415 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 981
- gi No. 2446981
- Description: (AB005560) AtGDI2 [Arabidopsis thaliana] >gi|2569936|emb|CAA04727| (AJ001397) GDI2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 416
- Location of Alignment in SEQ ID NO 2121: from 1 to 415

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2122
- Ceres seq_id 1571846
- Location of start within SEQ ID NO 2120: at 253 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GDP dissociation inhibitor
- Location within SEQ ID NO 2122: from 1 to 383 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 982
- gi No. 2446981
- Description: (AB005560) AtGDI2 [Arabidopsis thaliana] >gi|2569936|emb|CAA04727| (AJ001397) GDI2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 416
- Location of Alignment in SEQ ID NO 2122: from 1 to 383

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2123
- Ceres seq_id 1571847
- Location of start within SEQ ID NO 2120: at 379 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GDP dissociation inhibitor
- Location within SEQ ID NO 2123: from 1 to 341 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 983
- gi No. 2446981
- Description: (AB005560) AtGDI2 [Arabidopsis thaliana] >gi|2569936|emb|CAA04727| (AJ001397) GDI2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 416
- Location of Alignment in SEQ ID NO 2123: from 1 to 341

Maximum Length Sequence:
related to:
Clone IDs:
5188
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2124
- Ceres seq_id 1571867

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2125
- Ceres seq_id 1571868
- Location of start within SEQ ID NO 2124: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 2125: from 23 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 984
- gi No. 3335355
- Description: (AC004512) Match to polyubiquitin DNA gb|L05401 from A. thaliana. Contains insertion of mitochondrial NADH dehydrogenase gb|X82618 and gb|X98301. May be a pseudogene with an expressed insert. EST gb|AA586248 comes from this region. [Arabi...
- % Identity: 99.1
- Alignment Length: 228
- Location of Alignment in SEQ ID NO 2125: from 23 to 250

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2126
- Ceres seq_id 1571869
- Location of start within SEQ ID NO 2124: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 2126: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 985
- gi No. 3335355
- Description: (AC004512) Match to polyubiquitin DNA gb|L05401 from A. thaliana. Contains insertion of mitochondrial NADH dehydrogenase gb|X82618 and gb|X98301. May be a pseudogene with an expressed insert. EST gb|AA586248 comes from this region. [Arabi...
- % Identity: 99.1
- Alignment Length: 228
- Location of Alignment in SEQ ID NO 2126: from 1 to 228

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2127
- Ceres seq_id 1571870
- Location of start within SEQ ID NO 2124: at 297 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 2127: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 986
- gi No. 3335355
- Description: (AC004512) Match to polyubiquitin DNA gb|L05401 from A. thaliana. Contains insertion of mitochondrial NADH dehydrogenase gb|X82618 and gb|X98301. May be a pseudogene with an expressed insert. EST gb|AA586248 comes from this region. [Arabi...
- % Identity: 99.1
- Alignment Length: 228
- Location of Alignment in SEQ ID NO 2127: from 1 to 152

Maximum Length Sequence:
related to:
Clone IDs:

5256
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2128
- Ceres seq_id 1571871
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2129
- Ceres seq_id 1571872
- Location of start within SEQ ID NO 2128: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2129: from 1 to 185 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 987
- gi No. 118575
- Description: LATE EMBRYOGENESIS ABUNDANT PROTEIN (LEA PROTEIN) >gi|99846|pir||S12095 embryonic abundant protein (clone RSLEA2) - radish >gi|21120|emb|CAA39727| (X56280) late embryogenesis abundant protein [Raphanus sativus]
- % Identity: 74.1
- Alignment Length: 193
- Location of Alignment in SEQ ID NO 2129: from 1 to 185

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2130
- Ceres seq_id 1571873
- Location of start within SEQ ID NO 2128: at 182 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2130: from 1 to 153 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 988
- gi No. 118575
- Description: LATE EMBRYOGENESIS ABUNDANT PROTEIN (LEA PROTEIN) >gi|99846|pir||S12095 embryonic abundant protein (clone RSLEA2) - radish >gi|21120|emb|CAA39727| (X56280) late embryogenesis abundant protein [Raphanus sativus]
- % Identity: 74.1
- Alignment Length: 193
- Location of Alignment in SEQ ID NO 2130: from 1 to 153

Maximum Length Sequence:
related to:
Clone IDs:
533
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2131
- Ceres seq_id 1571874
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2132
- Ceres seq_id 1571875
- Location of start within SEQ ID NO 2131: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase

- Location within SEQ ID NO 2132: from 31 to 224 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2133
- Ceres seq_id 1571876
- Location of start within SEQ ID NO 2131: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2133: from 13 to 206 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2134
- Ceres seq_id 1571877
- Location of start within SEQ ID NO 2131: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2134: from 1 to 165 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
5485
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2135
- Ceres seq_id 1571884
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2136
- Ceres seq_id 1571885
- Location of start within SEQ ID NO 2135: at 16 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 989
- gi No. 3935155
- Description: (AC005106) T25N20.19 [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 2136: from 25 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2137
- Ceres seq_id 1571886
- Location of start within SEQ ID NO 2135: at 235 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 990
- gi No. 3935155
- Description: (AC005106) T25N20.19 [Arabidopsis thaliana]

- % Identity: 96.1
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 2137: from 1 to 78

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2138
- Ceres seq_id 1571887
- Location of start within SEQ ID NO 2135: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 991
- gi No. 3935155
- Description: (AC005106) T25N20.19 [Arabidopsis thaliana]
- % Identity: 96.1
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 2138: from 1 to 74

Maximum Length Sequence:
related to:
Clone IDs:
5507

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2139
- Ceres seq_id 1571888

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2140
- Ceres seq_id 1571889
- Location of start within SEQ ID NO 2139: at 362 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2140: from 1 to 183 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 992
- gi No. 99698
- Description: glutamate--ammonia ligase (EC 6.3.1.2), cytosolic (clone lambdaAtgskb6) - Arabidopsis thaliana
- % Identity: 96.9
- Alignment Length: 193
- Location of Alignment in SEQ ID NO 2140: from 1 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2141
- Ceres seq_id 1571890
- Location of start within SEQ ID NO 2139: at 494 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2141: from 1 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 993
- gi No. 99698
- Description: glutamate--ammonia ligase (EC 6.3.1.2), cytosolic (clone lambdaAtgskb6) - Arabidopsis thaliana
- % Identity: 96.9

- Alignment Length: 193
- Location of Alignment in SEQ ID NO 2141: from 1 to 139

Maximum Length Sequence:
related to:
Clone IDs:
5678
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2142
- Ceres seq_id 1571891
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2143
- Ceres seq_id 1571892
- Location of start within SEQ ID NO 2142: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 2143: from 41 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 994
- gi No. 121627
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR >gi|82244|pir||A26099 glycine-rich cell wall structural protein - garden petunia >gi|20553|emb|CAA27866| (X04335) precursor polypeptide (aa -27 to 357) [Petunia x hybrida]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2143: from 137 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2144
- Ceres seq_id 1571893
- Location of start within SEQ ID NO 2142: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 2144: from 25 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 995
- gi No. 121627
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR >gi|82244|pir||A26099 glycine-rich cell wall structural protein - garden petunia >gi|20553|emb|CAA27866| (X04335) precursor polypeptide (aa -27 to 357) [Petunia x hybrida]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2144: from 121 to 132

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2145
- Ceres seq_id 1571894
- Location of start within SEQ ID NO 2142: at 127 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 2145: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 996
- gi No. 121627
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR >gi|82244|pir||A26099 glycine-rich cell wall structural protein - garden petunia >gi|20553|emb|CAA27866| (X04335) precursor polypeptide (aa -27 to 357) [Petunia x hybrida]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2145: from 95 to 106

Maximum Length Sequence:
related to:
Clone IDs:
5710
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2146
- Ceres seq_id 1571898
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2147
- Ceres seq_id 1571899
- Location of start within SEQ ID NO 2146: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ferritins
- Location within SEQ ID NO 2147: from 92 to 249 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 997
- gi No. 2970654
- Description: (AF052058) ferritin subunit cowpea2 precursor [Vigna unguiculata]
- % Identity: 70.9
- Alignment Length: 244
- Location of Alignment in SEQ ID NO 2147: from 14 to 254

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2148
- Ceres seq_id 1571900
- Location of start within SEQ ID NO 2146: at 10 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ferritins
- Location within SEQ ID NO 2148: from 89 to 246 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 998
- gi No. 2970654
- Description: (AF052058) ferritin subunit cowpea2 precursor [Vigna unguiculata]
- % Identity: 70.9
- Alignment Length: 244
- Location of Alignment in SEQ ID NO 2148: from 11 to 251

Maximum Length Sequence:
related to:
Clone IDs:
6220

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2149
- Ceres seq_id 1571905

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2150
- Ceres seq_id 1571906
- Location of start within SEQ ID NO 2149: at 379 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 2150: from 21 to 243 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 999
- gi No. 1171771
- Description: NITRILASE 4 >gi|508737 (U09961) nitrilase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 235
- Location of Alignment in SEQ ID NO 2150: from 21 to 255

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2151
- Ceres seq_id 1571907
- Location of start within SEQ ID NO 2149: at 397 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 2151: from 15 to 237 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1000
- gi No. 1171771
- Description: NITRILASE 4 >gi|508737 (U09961) nitrilase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 235
- Location of Alignment in SEQ ID NO 2151: from 15 to 249

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2152
- Ceres seq_id 1571908
- Location of start within SEQ ID NO 2149: at 412 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nitrilase/cyanide hydratase
- Location within SEQ ID NO 2152: from 10 to 232 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1001
- gi No. 1171771
- Description: NITRILASE 4 >gi|508737 (U09961) nitrilase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 235
- Location of Alignment in SEQ ID NO 2152: from 10 to 244

Maximum Length Sequence:

related to:
Clone IDs:
6280
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2153
- Ceres seq_id 1571909
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2154
- Ceres seq_id 1571910
- Location of start within SEQ ID NO 2153: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1002
- gi No. 2558660
- Description: (AC002354) No definition line found [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 350
- Location of Alignment in SEQ ID NO 2154: from 1 to 350

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2155
- Ceres seq_id 1571911
- Location of start within SEQ ID NO 2153: at 183 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1003
- gi No. 2558660
- Description: (AC002354) No definition line found [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 350
- Location of Alignment in SEQ ID NO 2155: from 1 to 326

Maximum Length Sequence:
related to:
Clone IDs:
6397
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2156
- Ceres seq_id 1571912
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2157
- Ceres seq_id 1571913
- Location of start within SEQ ID NO 2156: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2157: from 89 to 154 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1004
- gi No. 2344900
- Description: (AC002388) EREBP isolog [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 226

- Location of Alignment in SEQ ID NO 2157: from 1 to 226

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2158
- Ceres seq_id 1571914
- Location of start within SEQ ID NO 2156: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2158: from 75 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1005
- gi No. 2344900
- Description: (AC002388) EREBP isolog [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 226
- Location of Alignment in SEQ ID NO 2158: from 1 to 212

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2159
- Ceres seq_id 1571915
- Location of start within SEQ ID NO 2156: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2159: from 46 to 111 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1006
- gi No. 2344900
- Description: (AC002388) EREBP isolog [Arabidopsis thaliana]
- % Identity: 98.7
- Alignment Length: 226
- Location of Alignment in SEQ ID NO 2159: from 1 to 183

Maximum Length Sequence:
related to:
Clone IDs:
6457
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2160
- Ceres seq_id 1571920
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2161
- Ceres seq_id 1571921
- Location of start within SEQ ID NO 2160: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- MSP (Major sperm protein) domain.
- Location within SEQ ID NO 2161: from 6 to 134 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2162
- Ceres seq_id 1571922
- Location of start within SEQ ID NO 2160: at 370 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- MSP (Major sperm protein) domain.
- Location within SEQ ID NO 2162: from 1 to 65 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2163
- Ceres seq_id 1571923
- Location of start within SEQ ID NO 2160: at 478 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6579
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2164
- Ceres seq_id 1571927
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2165
- Ceres seq_id 1571928
- Location of start within SEQ ID NO 2164: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridoxal-dependent decarboxylase
- Location within SEQ ID NO 2165: from 106 to 462 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2166
- Ceres seq_id 1571929
- Location of start within SEQ ID NO 2164: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridoxal-dependent decarboxylase
- Location within SEQ ID NO 2166: from 95 to 451 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2167
- Ceres seq_id 1571930
- Location of start within SEQ ID NO 2164: at 289 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pyridoxal-dependent decarboxylase
- Location within SEQ ID NO 2167: from 10 to 366 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

related to:
Clone IDs:
6596
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2168
- Ceres seq_id 1571931
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2169
- Ceres seq_id 1571932
- Location of start within SEQ ID NO 2168: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 2169: from 1 to 301 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1007
- gi No. 585421
- Description: LIPOXYGENASE, CHLOROPLAST PRECURSOR >gi|541879|pir||JQ2391 lipoxygenase (EC 1.13.11.12) AtLox2 - Arabidopsis thaliana >gi|431258 (L23968) lipoxygenase [Arabidopsis thaliana]
- % Identity: 93.9
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 2169: from 1 to 301

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2170
- Ceres seq_id 1571933
- Location of start within SEQ ID NO 2168: at 326 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 2170: from 1 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1008
- gi No. 585421
- Description: LIPOXYGENASE, CHLOROPLAST PRECURSOR >gi|541879|pir||JQ2391 lipoxygenase (EC 1.13.11.12) AtLox2 - Arabidopsis thaliana >gi|431258 (L23968) lipoxygenase [Arabidopsis thaliana]
- % Identity: 93.9
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 2170: from 1 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2171
- Ceres seq_id 1571934
- Location of start within SEQ ID NO 2168: at 371 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipoxygenase
- Location within SEQ ID NO 2171: from 1 to 178 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1009
- gi No. 585421

- Description: LIPOXYGENASE, CHLOROPLAST PRECURSOR
>gi|541879|pir||JQ2391 lipoxygenase (EC 1.13.11.12) AtLox2 - Arabidopsis thaliana >gi|431258 (L23968) lipoxygenase [Arabidopsis thaliana]
- % Identity: 93.9
- Alignment Length: 327
- Location of Alignment in SEQ ID NO 2171: from 1 to 178

Maximum Length Sequence:
related to:
Clone IDs:
6610
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2172
- Ceres seq_id 1571935
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2173
- Ceres seq_id 1571936
- Location of start within SEQ ID NO 2172: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2173: from 190 to 230 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2174
- Ceres seq_id 1571937
- Location of start within SEQ ID NO 2172: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2174: from 152 to 192 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2175
- Ceres seq_id 1571938
- Location of start within SEQ ID NO 2172: at 389 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2175: from 144 to 184 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6757
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2176
- Ceres seq_id 1571945
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2177
- Ceres seq_id 1571946
- Location of start within SEQ ID NO 2176: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1010
- gi No. 1709794
- Description: 26S PROTEASOME REGULATORY SUBUNIT S5A (MULTIUBIQUITIN CHAIN BINDING PROTEIN) >gi|1165206 (U33269) MBP1 [Arabidopsis thaliana] >gi|4467150|emb|CAB37519| (AL035540) multiubiquitin chain binding protein (MBP1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 386
- Location of Alignment in SEQ ID NO 2177: from 1 to 386

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2178
- Ceres seq_id 1571947
- Location of start within SEQ ID NO 2176: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1011
- gi No. 1709794
- Description: 26S PROTEASOME REGULATORY SUBUNIT S5A (MULTIUBIQUITIN CHAIN BINDING PROTEIN) >gi|1165206 (U33269) MBP1 [Arabidopsis thaliana] >gi|4467150|emb|CAB37519| (AL035540) multiubiquitin chain binding protein (MBP1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 386
- Location of Alignment in SEQ ID NO 2178: from 1 to 380

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2179
- Ceres seq_id 1571948
- Location of start within SEQ ID NO 2176: at 127 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1012
- gi No. 1709794
- Description: 26S PROTEASOME REGULATORY SUBUNIT S5A (MULTIUBIQUITIN CHAIN BINDING PROTEIN) >gi|1165206 (U33269) MBP1 [Arabidopsis thaliana] >gi|4467150|emb|CAB37519| (AL035540) multiubiquitin chain binding protein (MBP1) [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 386
- Location of Alignment in SEQ ID NO 2179: from 1 to 371

Maximum Length Sequence:
related to:
Clone IDs:
6789

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2180
- Ceres seq_id 1571949

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2181
- Ceres seq_id 1571950
- Location of start within SEQ ID NO 2180: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1013
- gi No. 642484
- Description: (U16371) androgen receptor [Homo sapiens]
- % Identity: 70.4
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 2181: from 78 to 104

Maximum Length Sequence:
related to:
Clone IDs:
7101
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2182
- Ceres seq_id 1571963
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2183
- Ceres seq_id 1571964
- Location of start within SEQ ID NO 2182: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HIT family
- Location within SEQ ID NO 2183: from 8 to 105 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2184
- Ceres seq_id 1571965
- Location of start within SEQ ID NO 2182: at 284 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
7434
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2185
- Ceres seq_id 1571987
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2186
- Ceres seq_id 1571988
- Location of start within SEQ ID NO 2185: at 217 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2186: from 70 to 242 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1014
- gi No. 2586125
- Description: (U89512) b-keto acyl reductase [Arabidopsis thaliana]

- % Identity: 99.6
- Alignment Length: 249
- Location of Alignment in SEQ ID NO 2186: from 70 to 318

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2187
- Ceres seq_id 1571989
- Location of start within SEQ ID NO 2185: at 541 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2187: from 1 to 134 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1015
- gi No. 2586125
- Description: (U89512) b-keto acyl reductase [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 249
- Location of Alignment in SEQ ID NO 2187: from 1 to 210

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2188
- Ceres seq_id 1571990
- Location of start within SEQ ID NO 2185: at 634 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2188: from 1 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1016
- gi No. 2586125
- Description: (U89512) b-keto acyl reductase [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 249
- Location of Alignment in SEQ ID NO 2188: from 1 to 179

Maximum Length Sequence:
related to:
Clone IDs:
7700
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2189
- Ceres seq_id 1571999

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2190
- Ceres seq_id 1572000
- Location of start within SEQ ID NO 2189: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 2190: from 54 to 217 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1017

- gi No. 4741950
- Description: (AF134125) Lhcb2 protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 266
- Location of Alignment in SEQ ID NO 2190: from 17 to 217

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2191
- Ceres seq_id 1572001
- Location of start within SEQ ID NO 2189: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 2191: from 38 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1018
- gi No. 4741950
- Description: (AF134125) Lhcb2 protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 266
- Location of Alignment in SEQ ID NO 2191: from 1 to 201

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2192
- Ceres seq_id 1572002
- Location of start within SEQ ID NO 2189: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 2192: from 1 to 164 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1019
- gi No. 4741950
- Description: (AF134125) Lhcb2 protein [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 266
- Location of Alignment in SEQ ID NO 2192: from 1 to 164

Maximum Length Sequence:
related to:
Clone IDs:
819
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2193
- Ceres seq_id 1572023

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2194
- Ceres seq_id 1572024
- Location of start within SEQ ID NO 2193: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 2194: from 89 to 154 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2195
- Ceres seq_id 1572025
- Location of start within SEQ ID NO 2193: at 274 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 2195: from 19 to 84 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
8381
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2196
- Ceres seq_id 1572026
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2197
- Ceres seq_id 1572027
- Location of start within SEQ ID NO 2196: at 193 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1020
- gi No. 1084337
- Description: drought-induced protein Di21 - Arabidopsis thaliana >gi|469112|emb|CAA55322| (X78585) Di21 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 104
- Location of Alignment in SEQ ID NO 2197: from 1 to 104

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2198
- Ceres seq_id 1572028
- Location of start within SEQ ID NO 2196: at 346 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1021
- gi No. 1084337
- Description: drought-induced protein Di21 - Arabidopsis thaliana >gi|469112|emb|CAA55322| (X78585) Di21 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 104
- Location of Alignment in SEQ ID NO 2198: from 1 to 53

Maximum Length Sequence:
related to:
Clone IDs:
8793
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2199
- Ceres seq_id 1572029
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2200
- Ceres seq_id 1572030

- Location of start within SEQ ID NO 2199: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Barwin family
- Location within SEQ ID NO 2200: from 83 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1022
- gi No. 1170247
- Description: HEVEIN-LIKE PROTEIN PRECURSOR >gi|407248 (U01880) pre-hevein-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 2200: from 11 to 222

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2201
- Ceres seq_id 1572031
- Location of start within SEQ ID NO 2199: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Barwin family
- Location within SEQ ID NO 2201: from 73 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1023
- gi No. 1170247
- Description: HEVEIN-LIKE PROTEIN PRECURSOR >gi|407248 (U01880) pre-hevein-like protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 2201: from 1 to 212

Maximum Length Sequence:
related to:
Clone IDs:
8992
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2202
- Ceres seq_id 1572040
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2203
- Ceres seq_id 1572041
- Location of start within SEQ ID NO 2202: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2203: from 73 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1024
- gi No. 1079078
- Description: GCR 101 protein - fruit fly (Drosophila melanogaster) >gi|510509|emb|CAA50795| (X71975) put. homologue to S.cerevisiae GAR1 gene [Drosophila melanogaster]
- % Identity: 70.2
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 2203: from 73 to 118

Maximum Length Sequence:
related to:
Clone IDs:
9012
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2204
- Ceres seq_id 1572046
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2205
- Ceres seq_id 1572047
- Location of start within SEQ ID NO 2204: at 349 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1025
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 2205: from 65 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2206
- Ceres seq_id 1572048
- Location of start within SEQ ID NO 2204: at 472 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1026
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 2206: from 24 to 56

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2207
- Ceres seq_id 1572049
- Location of start within SEQ ID NO 2204: at 475 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1027
- gi No. 2130118
- Description: amylogenin - maize (fragments)
- % Identity: 81.8
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 2207: from 23 to 55

Maximum Length Sequence:
related to:
Clone IDs:
954
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2208
- Ceres seq_id 1572064
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2209
- Ceres seq_id 1572065
- Location of start within SEQ ID NO 2208: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1028
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2209: from 1 to 141

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2210
- Ceres seq_id 1572066
- Location of start within SEQ ID NO 2208: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1029
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2210: from 1 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2211
- Ceres seq_id 1572067
- Location of start within SEQ ID NO 2208: at 141 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1030
- gi No. 2494128
- Description: (AC002376) EST gb|T43244 comes from this gene. [Arabidopsis thaliana]
- % Identity: 79.3
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2211: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
9897
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2212
- Ceres seq_id 1572082
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2213
- Ceres seq_id 1572083
- Location of start within SEQ ID NO 2212: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Ribosomal protein L14
- Location within SEQ ID NO 2213: from 19 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1031
- gi No. 730536
- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 2213: from 1 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2214
- Ceres seq_id 1572084
- Location of start within SEQ ID NO 2212: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 2214: from 4 to 125 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1032
- gi No. 730536
- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 2214: from 1 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2215
- Ceres seq_id 1572085
- Location of start within SEQ ID NO 2212: at 241 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 2215: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1033
- gi No. 730536
- Description: 60S RIBOSOMAL PROTEIN L23 >gi|310933 (L18915) 60S ribosomal protein subunit L17 [Nicotiana tabacum]
- % Identity: 95
- Alignment Length: 140
- Location of Alignment in SEQ ID NO 2215: from 1 to 81

Maximum Length Sequence:
related to:
Clone IDs:
43057
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2216
- Ceres seq_id 1572094

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2217
- Ceres seq_id 1572095
- Location of start within SEQ ID NO 2216: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2217: from 53 to 122 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2218
- Ceres seq_id 1572096
- Location of start within SEQ ID NO 2216: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2218: from 27 to 96 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
42101
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2219
- Ceres seq_id 1572097
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2220
- Ceres seq_id 1572098
- Location of start within SEQ ID NO 2219: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adhesion lipoprotein
- Location within SEQ ID NO 2220: from 118 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1034
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]
- % Identity: 75.7
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 2220: from 1 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2221
- Ceres seq_id 1572099
- Location of start within SEQ ID NO 2219: at 336 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adhesion lipoprotein
- Location within SEQ ID NO 2221: from 27 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1035
- gi No. 2244762
- Description: (Z97335) major latex protein like [Arabidopsis thaliana]

- % Identity: 75.7
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 2221: from 1 to 57

Maximum Length Sequence:
related to:
Clone IDs:
232091
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2222
- Ceres seq_id 1572120
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2223
- Ceres seq_id 1572121
- Location of start within SEQ ID NO 2222: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1036
- gi No. 3047065
- Description: (AF058825) contains similarity to human OS-9 precurosor (GB:U41635) [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 88
- Location of Alignment in SEQ ID NO 2223: from 64 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2224
- Ceres seq_id 1572122
- Location of start within SEQ ID NO 2222: at 264 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1037
- gi No. 3047065
- Description: (AF058825) contains similarity to human OS-9 precurosor (GB:U41635) [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 88
- Location of Alignment in SEQ ID NO 2224: from 1 to 88

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2225
- Ceres seq_id 1572123
- Location of start within SEQ ID NO 2222: at 270 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1038
- gi No. 3047065
- Description: (AF058825) contains similarity to human OS-9 precurosor (GB:U41635) [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 88
- Location of Alignment in SEQ ID NO 2225: from 1 to 86

Maximum Length Sequence:
related to:

Clone IDs:
232402
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2226
- Ceres seq_id 1572139
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2227
- Ceres seq_id 1572140
- Location of start within SEQ ID NO 2226: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2227: from 41 to 490 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2228
- Ceres seq_id 1572141
- Location of start within SEQ ID NO 2226: at 42 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2228: from 28 to 477 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2229
- Ceres seq_id 1572142
- Location of start within SEQ ID NO 2226: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2229: from 12 to 461 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
247831
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2230
- Ceres seq_id 1572147
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2231
- Ceres seq_id 1572148
- Location of start within SEQ ID NO 2230: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 2231: from 24 to 382 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1039
- gi No. 1169598

- Description: OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM (DELTA-12 DESATURASE) >gi|438451 (L26296) delta-12 desaturase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 383
- Location of Alignment in SEQ ID NO 2231: from 1 to 383

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2232
- Ceres seq_id 1572149
- Location of start within SEQ ID NO 2230: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 2232: from 18 to 376 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1040
- gi No. 1169598
- Description: OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM (DELTA-12 DESATURASE) >gi|438451 (L26296) delta-12 desaturase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 383
- Location of Alignment in SEQ ID NO 2232: from 1 to 377

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2233
- Ceres seq_id 1572150
- Location of start within SEQ ID NO 2230: at 675 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Fatty acid desaturase
- Location within SEQ ID NO 2233: from 1 to 203 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1041
- gi No. 1169598
- Description: OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM (DELTA-12 DESATURASE) >gi|438451 (L26296) delta-12 desaturase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 383
- Location of Alignment in SEQ ID NO 2233: from 1 to 204

Maximum Length Sequence:
related to:
Clone IDs:
249844
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2234
- Ceres seq_id 1572174

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2235
- Ceres seq_id 1572175
- Location of start within SEQ ID NO 2234: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Tropomyosins
- Location within SEQ ID NO 2235: from 122 to 242 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
250383
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2236
- Ceres seq_id 1572180
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2237
- Ceres seq_id 1572181
- Location of start within SEQ ID NO 2236: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S9/S16
- Location within SEQ ID NO 2237: from 87 to 208 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1042
- gi No. 5456946
- Description: (AB022676) ribosomal protein S9 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 208
- Location of Alignment in SEQ ID NO 2237: from 1 to 208

Maximum Length Sequence:
related to:
Clone IDs:
251247
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2238
- Ceres seq_id 1572195
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2239
- Ceres seq_id 1572196
- Location of start within SEQ ID NO 2238: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Response regulator receiver domain
- Location within SEQ ID NO 2239: from 22 to 136 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1043
- gi No. 3687688
- Description: (AF057027) response regulator protein [Brassica napus]
- % Identity: 78.9
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 2239: from 1 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2240
- Ceres seq_id 1572197
- Location of start within SEQ ID NO 2238: at 210 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Response regulator receiver domain
- Location within SEQ ID NO 2240: from 1 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1044
- gi No. 3687688
- Description: (AF057027) response regulator protein [Brassica napus]
- % Identity: 78.9
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 2240: from 1 to 104

Maximum Length Sequence:
related to:
Clone IDs:
251466
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2241
- Ceres seq_id 1572201
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2242
- Ceres seq_id 1572202
- Location of start within SEQ ID NO 2241: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2242: from 62 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1045
- gi No. 1246403
- Description: (X94698) TINY [Arabidopsis thaliana] >gi|3406035 (AC005405) TINY [Arabidopsis thaliana]
- % Identity: 74.5
- Alignment Length: 55
- Location of Alignment in SEQ ID NO 2242: from 59 to 112

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2243
- Ceres seq_id 1572203
- Location of start within SEQ ID NO 2241: at 490 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
251549
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2244
- Ceres seq_id 1572204
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2245
- Ceres seq_id 1572205
- Location of start within SEQ ID NO 2244: at 430 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1046
- gi No. 4982498
- Description: (AC000107) F17F8.4 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2245: from 6 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2246
- Ceres seq_id 1572206
- Location of start within SEQ ID NO 2244: at 439 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1047
- gi No. 4982498
- Description: (AC000107) F17F8.4 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2246: from 3 to 156

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2247
- Ceres seq_id 1572207
- Location of start within SEQ ID NO 2244: at 466 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1048
- gi No. 4982498
- Description: (AC000107) F17F8.4 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2247: from 1 to 147

Maximum Length Sequence:
related to:
Clone IDs:
248156
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2248
- Ceres seq_id 1572240
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2249
- Ceres seq_id 1572241
- Location of start within SEQ ID NO 2248: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 2249: from 38 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1049
- gi No. 4090200

- Description: (AJ012310) WUSCHEL protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 2249: from 1 to 292

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2250
- Ceres seq_id 1572242
- Location of start within SEQ ID NO 2248: at 441 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1050
- gi No. 4090200
- Description: (AJ012310) WUSCHEL protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 2250: from 1 to 187

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2251
- Ceres seq_id 1572243
- Location of start within SEQ ID NO 2248: at 474 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1051
- gi No. 4090200
- Description: (AJ012310) WUSCHEL protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 292
- Location of Alignment in SEQ ID NO 2251: from 1 to 176

Maximum Length Sequence:
related to:
Clone IDs:
248969
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2252
- Ceres seq_id 1572244

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2253
- Ceres seq_id 1572245
- Location of start within SEQ ID NO 2252: at 184 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2254
- Ceres seq_id 1572246
- Location of start within SEQ ID NO 2252: at 320 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2254: from 63 to 115 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
251841
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2255
- Ceres seq_id 1572298
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2256
- Ceres seq_id 1572299
- Location of start within SEQ ID NO 2255: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2256: from 46 to 113 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2257
- Ceres seq_id 1572300
- Location of start within SEQ ID NO 2255: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2257: from 4 to 71 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2258
- Ceres seq_id 1572301
- Location of start within SEQ ID NO 2255: at 249 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
254278
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2259
- Ceres seq_id 1572338
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2260
- Ceres seq_id 1572339
- Location of start within SEQ ID NO 2259: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 2260: from 212 to 250 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2261
- Ceres seq_id 1572340
- Location of start within SEQ ID NO 2259: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 2261: from 191 to 229 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2262
- Ceres seq_id 1572341
- Location of start within SEQ ID NO 2259: at 295 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 2262: from 114 to 152 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
254633
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2263
- Ceres seq_id 1572349

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2264
- Ceres seq_id 1572350
- Location of start within SEQ ID NO 2263: at 98 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1052
- gi No. 2935529
- Description: (AF049069) No definition line found [Pinus radiata]
- % Identity: 76
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 2264: from 101 to 317

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2265
- Ceres seq_id 1572351
- Location of start within SEQ ID NO 2263: at 242 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1053
- gi No. 2935529
- Description: (AF049069) No definition line found [Pinus radiata]
- % Identity: 76
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 2265: from 53 to 269

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2266
- Ceres seq_id 1572352
- Location of start within SEQ ID NO 2263: at 323 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1054
- gi No. 2935529
- Description: (AF049069) No definition line found [Pinus radiata]
- % Identity: 76
- Alignment Length: 217
- Location of Alignment in SEQ ID NO 2266: from 26 to 242

Maximum Length Sequence:
related to:
Clone IDs:
255294
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2267
- Ceres seq_id 1572364
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2268
- Ceres seq_id 1572365
- Location of start within SEQ ID NO 2267: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 2268: from 48 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1055
- gi No. 4185515
- Description: (AF102824) actin depolymerizing factor 6 [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2268: from 30 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2269
- Ceres seq_id 1572366
- Location of start within SEQ ID NO 2267: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 2269: from 19 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1056
- gi No. 4185515
- Description: (AF102824) actin depolymerizing factor 6 [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2269: from 1 to 146

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2270
- Ceres seq_id 1572367
- Location of start within SEQ ID NO 2267: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 2270: from 5 to 132 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1057
- gi No. 4185515
- Description: (AF102824) actin depolymerizing factor 6 [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2270: from 1 to 132

Maximum Length Sequence:
related to:
Clone IDs:
255685
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2271
- Ceres seq_id 1572368
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2272
- Ceres seq_id 1572369
- Location of start within SEQ ID NO 2271: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2272: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2273
- Ceres seq_id 1572370
- Location of start within SEQ ID NO 2271: at 434 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
256067
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2274
- Ceres seq_id 1572375
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2275
- Ceres seq_id 1572376
- Location of start within SEQ ID NO 2274: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family

- Location within SEQ ID NO 2275: from 36 to 294 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2276
- Ceres seq_id 1572377
- Location of start within SEQ ID NO 2274: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 2276: from 30 to 288 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2277
- Ceres seq_id 1572378
- Location of start within SEQ ID NO 2274: at 315 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 2277: from 1 to 190 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
257992
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2278
- Ceres seq_id 1572397

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2279
- Ceres seq_id 1572398
- Location of start within SEQ ID NO 2278: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pathogenesis-related protein Bet v I family
- Location within SEQ ID NO 2279: from 12 to 156 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2280
- Ceres seq_id 1572399
- Location of start within SEQ ID NO 2278: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pathogenesis-related protein Bet v I family
- Location within SEQ ID NO 2280: from 1 to 130 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
258180
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2281
- Ceres seq_id 1572400
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2282
- Ceres seq_id 1572401
- Location of start within SEQ ID NO 2281: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2283
- Ceres seq_id 1572402
- Location of start within SEQ ID NO 2281: at 299 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1058
- gi No. 1169840
- Description: GIBBERELLIN-REGULATED PROTEIN 4 PRECURSOR >gi|2129591|pir||S71442 GAST1 protein homolog (clone GASA4) - Arabidopsis thaliana >gi|950099 (U15683) gibberellin-regulated [Arabidopsis thaliana]
- % Identity: 75.4
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 2283: from 1 to 41

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2284
- Ceres seq_id 1572403
- Location of start within SEQ ID NO 2281: at 328 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
258270
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2285
- Ceres seq_id 1572404
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2286
- Ceres seq_id 1572405
- Location of start within SEQ ID NO 2285: at 344 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 2286: from 249 to 371 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2287

- Ceres seq_id 1572406
- Location of start within SEQ ID NO 2285: at 347 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 2287: from 248 to 370 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2288
- Ceres seq_id 1572407
- Location of start within SEQ ID NO 2285: at 350 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family UPF0003
- Location within SEQ ID NO 2288: from 247 to 369 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
265926
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2289
- Ceres seq_id 1572448
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2290
- Ceres seq_id 1572449
- Location of start within SEQ ID NO 2289: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1059
- gi No. 2652938
- Description: (Z47554) orf [Zea mays]
- % Identity: 80.6
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 2290: from 6 to 36

Maximum Length Sequence:
related to:
Clone IDs:
266518
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2291
- Ceres seq_id 1572450
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2292
- Ceres seq_id 1572451
- Location of start within SEQ ID NO 2291: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2292: from 72 to 264 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1060
- gi No. 957251
- Description: (S77418) oxidoreductase required for shoot apex development=FEY [Arabidopsis thaliana, Peptide, 355 aa] [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 2292: from 72 to 376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2293
- Ceres seq_id 1572452
- Location of start within SEQ ID NO 2291: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2293: from 50 to 242 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1061
- gi No. 957251
- Description: (S77418) oxidoreductase required for shoot apex development=FEY [Arabidopsis thaliana, Peptide, 355 aa] [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 2293: from 50 to 354

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2294
- Ceres seq_id 1572453
- Location of start within SEQ ID NO 2291: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 2294: from 47 to 239 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1062
- gi No. 957251
- Description: (S77418) oxidoreductase required for shoot apex development=FEY [Arabidopsis thaliana, Peptide, 355 aa] [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 305
- Location of Alignment in SEQ ID NO 2294: from 47 to 351

Maximum Length Sequence:
related to:
Clone IDs:
266771
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2295
- Ceres seq_id 1572458
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2296
- Ceres seq_id 1572459
- Location of start within SEQ ID NO 2295: at 450 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 2296: from 57 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1063
- gi No. 3551247
- Description: (AB012703) 181 [Daucus carota]
- % Identity: 93
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2296: from 3 to 202

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2297
- Ceres seq_id 1572460
- Location of start within SEQ ID NO 2295: at 645 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1064
- gi No. 3551247
- Description: (AB012703) 181 [Daucus carota]
- % Identity: 93
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2297: from 1 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2298
- Ceres seq_id 1572461
- Location of start within SEQ ID NO 2295: at 675 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1065
- gi No. 3551247
- Description: (AB012703) 181 [Daucus carota]
- % Identity: 93
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2298: from 1 to 127

Maximum Length Sequence:
related to:
Clone IDs:
266884
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2299
- Ceres seq_id 1572462
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2300
- Ceres seq_id 1572463
- Location of start within SEQ ID NO 2299: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 2300: from 69 to 114 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2301
- Ceres seq_id 1572464
- Location of start within SEQ ID NO 2299: at 44 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 2301: from 55 to 100 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2302
- Ceres seq_id 1572465
- Location of start within SEQ ID NO 2299: at 345 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
268047
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2303
- Ceres seq_id 1572487
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2304
- Ceres seq_id 1572488
- Location of start within SEQ ID NO 2303: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2304: from 95 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1066
- gi No. 5091556
- Description: (AC007067) T10O24.25 [Arabidopsis thaliana]
- % Identity: 73.8
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 2304: from 12 to 183

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2305
- Ceres seq_id 1572489
- Location of start within SEQ ID NO 2303: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2305: from 84 to 164 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1067
- gi No. 5091556

- Description: (AC007067) T10O24.25 [Arabidopsis thaliana]
- % Identity: 73.8
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 2305: from 1 to 172

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2306
- Ceres seq_id 1572490
- Location of start within SEQ ID NO 2303: at 350 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2306: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1068
- gi No. 5091556
- Description: (AC007067) T10O24.25 [Arabidopsis thaliana]
- % Identity: 73.8
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 2306: from 1 to 67

Maximum Length Sequence:
related to:
Clone IDs:
268536

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2307
- Ceres seq_id 1572495

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2308
- Ceres seq_id 1572496
- Location of start within SEQ ID NO 2307: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2308: from 117 to 222 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2309
- Ceres seq_id 1572497
- Location of start within SEQ ID NO 2307: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2309: from 105 to 210 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2310
- Ceres seq_id 1572498
- Location of start within SEQ ID NO 2307: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2310: from 23 to 128 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
268795
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2311
- Ceres seq_id 1572503
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2312
- Ceres seq_id 1572504
- Location of start within SEQ ID NO 2311: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2312: from 12 to 89 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2313
- Ceres seq_id 1572505
- Location of start within SEQ ID NO 2311: at 25 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2313: from 4 to 81 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2314
- Ceres seq_id 1572506
- Location of start within SEQ ID NO 2311: at 392 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
270281
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2315
- Ceres seq_id 1572524
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2316
- Ceres seq_id 1572525
- Location of start within SEQ ID NO 2315: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp70 protein
- Location within SEQ ID NO 2316: from 1 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1069
- gi No. 3157925
- Description: (AC002131) ESTs gb|F13979, gb|AA042648, gb|H77240, gb|T76189, gb|H35952, gb|T22022 and gb|T42279 come from this gene. A similar ORF is present on the opposite strand. [Arabidopsis thaliana]
- % Identity: 84
- Alignment Length: 100
- Location of Alignment in SEQ ID NO 2316: from 7 to 104

Maximum Length Sequence:
related to:
Clone IDs:
253770
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2317
- Ceres seq_id 1572526
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2318
- Ceres seq_id 1572527
- Location of start within SEQ ID NO 2317: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Inositol monophosphatase family
- Location within SEQ ID NO 2318: from 122 to 327 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1070
- gi No. 1354510
- Description: (U55205) HAL2-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 373
- Location of Alignment in SEQ ID NO 2318: from 1 to 373

Maximum Length Sequence:
related to:
Clone IDs:
256433
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2319
- Ceres seq_id 1572528
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2320
- Ceres seq_id 1572529
- Location of start within SEQ ID NO 2319: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 2320: from 82 to 185 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2321
- Ceres seq_id 1572530
- Location of start within SEQ ID NO 2319: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 2321: from 54 to 157 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2322
- Ceres seq_id 1572531
- Location of start within SEQ ID NO 2319: at 662 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Iron/Ascorbate oxidoreductase family
- Location within SEQ ID NO 2322: from 1 to 126 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
269483
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2323
- Ceres seq_id 1572538
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2324
- Ceres seq_id 1572539
- Location of start within SEQ ID NO 2323: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2324: from 59 to 221 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1071
- gi No. 2342726
- Description: (AC002341) peroxidase isolog [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 215
- Location of Alignment in SEQ ID NO 2324: from 7 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2325
- Ceres seq_id 1572540
- Location of start within SEQ ID NO 2323: at 19 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2325: from 53 to 215 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1072
- gi No. 2342726
- Description: (AC002341) peroxidase isolog [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 215
- Location of Alignment in SEQ ID NO 2325: from 1 to 215

Maximum Length Sequence:
related to:
Clone IDs:
254801
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2326
- Ceres seq_id 1572569
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2327
- Ceres seq_id 1572570
- Location of start within SEQ ID NO 2326: at 401 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 2327: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1073
- gi No. 170352
- Description: (M74101) hexameric polyubiquitin [Nicotiana sylvestris] >gi|870792(L05361) polyubiquitin [Arabidopsis thaliana] >gi|4115333 (L81139)ubiquitin [Pisum sativum] >gi|4115335 (L81140) ubiquitin [Pisumsativum]
- % Identity: 100
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2327: from 1 to 146

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2328
- Ceres seq_id 1572571
- Location of start within SEQ ID NO 2326: at 629 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 2328: from 1 to 70 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1074
- gi No. 170352
- Description: (M74101) hexameric polyubiquitin [Nicotiana sylvestris] >gi|870792(L05361) polyubiquitin [Arabidopsis thaliana] >gi|4115333 (L81139)ubiquitin [Pisum sativum] >gi|4115335 (L81140) ubiquitin [Pisumsativum]
- % Identity: 100
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2328: from 1 to 70

Maximum Length Sequence:
related to:
Clone IDs:
255542
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2329
- Ceres seq_id 1572576
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2330
- Ceres seq_id 1572577
- Location of start within SEQ ID NO 2329: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1075
- gi No. 5302801
- Description: (Z97342) glycoprotein homolog [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 473
- Location of Alignment in SEQ ID NO 2330: from 1 to 473

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2331
- Ceres seq_id 1572578
- Location of start within SEQ ID NO 2329: at 492 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1076
- gi No. 5302801
- Description: (Z97342) glycoprotein homolog [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 473
- Location of Alignment in SEQ ID NO 2331: from 1 to 326

Maximum Length Sequence:
related to:
Clone IDs:
260870

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2332
- Ceres seq_id 1572627

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2333
- Ceres seq_id 1572628
- Location of start within SEQ ID NO 2332: at 127 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2333: from 15 to 65 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2334
- Ceres seq_id 1572629
- Location of start within SEQ ID NO 2332: at 366 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2335
- Ceres seq_id 1572630
- Location of start within SEQ ID NO 2332: at 571 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
262460
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2336
- Ceres seq_id 1572643
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2337
- Ceres seq_id 1572644
- Location of start within SEQ ID NO 2336: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 2337: from 19 to 66 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1077
- gi No. 2280528
- Description: (AB005888) ATMYB3 [Arabidopsis thaliana]
- % Identity: 71.2
- Alignment Length: 229
- Location of Alignment in SEQ ID NO 2337: from 1 to 214

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2338
- Ceres seq_id 1572645
- Location of start within SEQ ID NO 2336: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 2338: from 1 to 43 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1078
- gi No. 2280528
- Description: (AB005888) ATMYB3 [Arabidopsis thaliana]
- % Identity: 71.2
- Alignment Length: 229
- Location of Alignment in SEQ ID NO 2338: from 1 to 191

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2339
- Ceres seq_id 1572646
- Location of start within SEQ ID NO 2336: at 359 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1079
- gi No. 2280528
- Description: (AB005888) ATMYB3 [Arabidopsis thaliana]
- % Identity: 71.2
- Alignment Length: 229
- Location of Alignment in SEQ ID NO 2339: from 1 to 131

Maximum Length Sequence:

related to:
Clone IDs:
263816
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2340
- Ceres seq_id 1572647
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2341
- Ceres seq_id 1572648
- Location of start within SEQ ID NO 2340: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2341: from 51 to 308 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1080
- gi No. 3450842
- Description: (AF080436) mitogen activated protein kinase kinase [Oryza sativa]
- % Identity: 76.9
- Alignment Length: 277
- Location of Alignment in SEQ ID NO 2341: from 37 to 313

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2342
- Ceres seq_id 1572649
- Location of start within SEQ ID NO 2340: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2342: from 24 to 281 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1081
- gi No. 3450842
- Description: (AF080436) mitogen activated protein kinase kinase [Oryza sativa]
- % Identity: 76.9
- Alignment Length: 277
- Location of Alignment in SEQ ID NO 2342: from 10 to 286

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2343
- Ceres seq_id 1572650
- Location of start within SEQ ID NO 2340: at 339 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2343: from 1 to 196 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1082
- gi No. 3450842
- Description: (AF080436) mitogen activated protein kinase kinase [Oryza sativa]
- % Identity: 76.9
- Alignment Length: 277

- Location of Alignment in SEQ ID NO 2343: from 1 to 201

Maximum Length Sequence:
related to:
Clone IDs:
266730
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2344
- Ceres seq_id 1572662
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2345
- Ceres seq_id 1572663
- Location of start within SEQ ID NO 2344: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2346
- Ceres seq_id 1572664
- Location of start within SEQ ID NO 2344: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2347
- Ceres seq_id 1572665
- Location of start within SEQ ID NO 2344: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1083
- gi No. 1513292
- Description: (U66523) ORFveg113; similar to lysozyme C, Swiss-Prot Accession Number P37712 [Dictyostelium discoideum]
- % Identity: 71.4
- Alignment Length: 42
- Location of Alignment in SEQ ID NO 2347: from 9 to 49

Maximum Length Sequence:
related to:
Clone IDs:
20948
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2348
- Ceres seq_id 1572694
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2349
- Ceres seq_id 1572695
- Location of start within SEQ ID NO 2348: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1084
- gi No. 100219

- Description: glycine-rich protein (clone uK-4) - tomato >gi|1345534|emb|CAA39225| (X55696) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 71.4
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 2349: from 5 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2350
- Ceres seq_id 1572696
- Location of start within SEQ ID NO 2348: at 185 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2351
- Ceres seq_id 1572697
- Location of start within SEQ ID NO 2348: at 254 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
270276
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2352
- Ceres seq_id 1572706
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2353
- Ceres seq_id 1572707
- Location of start within SEQ ID NO 2352: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2353: from 37 to 116 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1085
- gi No. 3892704
- Description: (AL033545) RCc3-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 2353: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
19244
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2354
- Ceres seq_id 1572750
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2355
- Ceres seq_id 1572751
- Location of start within SEQ ID NO 2354: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2355: from 134 to 195 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1086
- gi No. 2281637
- Description: (AF003099) AP2 domain containing protein RAP2.6 [Arabidopsis thaliana]
- % Identity: 83.1
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 2355: from 131 to 195

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2356
- Ceres seq_id 1572752
- Location of start within SEQ ID NO 2354: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2356: from 75 to 136 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1087
- gi No. 2281637
- Description: (AF003099) AP2 domain containing protein RAP2.6 [Arabidopsis thaliana]
- % Identity: 83.1
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 2356: from 72 to 136

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2357
- Ceres seq_id 1572753
- Location of start within SEQ ID NO 2354: at 187 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2357: from 72 to 133 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1088
- gi No. 2281637
- Description: (AF003099) AP2 domain containing protein RAP2.6 [Arabidopsis thaliana]
- % Identity: 83.1
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 2357: from 69 to 133

Maximum Length Sequence:
related to:
Clone IDs:
17347
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2358
- Ceres seq_id 1572762
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2359
- Ceres seq_id 1572763
- Location of start within SEQ ID NO 2358: at 552 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2359: from 30 to 111 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1089
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 98.6
- Alignment Length: 139
- Location of Alignment in SEQ ID NO 2359: from 1 to 114

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2360
- Ceres seq_id 1572764
- Location of start within SEQ ID NO 2358: at 573 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2360: from 23 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1090
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 98.6
- Alignment Length: 139
- Location of Alignment in SEQ ID NO 2360: from 1 to 107

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2361
- Ceres seq_id 1572765
- Location of start within SEQ ID NO 2358: at 592 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
28545

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2362
- Ceres seq_id 1572774

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2363
- Ceres seq_id 1572775
- Location of start within SEQ ID NO 2362: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L21e
- Location within SEQ ID NO 2363: from 16 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1091
- gi No. 3885884
- Description: (AF093630) 60S ribosomal protein L21 [Oryza sativa]
- % Identity: 87.2
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 2363: from 15 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2364
- Ceres seq_id 1572776
- Location of start within SEQ ID NO 2362: at 45 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L21e
- Location within SEQ ID NO 2364: from 2 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1092
- gi No. 3885884
- Description: (AF093630) 60S ribosomal protein L21 [Oryza sativa]
- % Identity: 87.2
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 2364: from 1 to 164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2365
- Ceres seq_id 1572777
- Location of start within SEQ ID NO 2362: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1093
- gi No. 3885884
- Description: (AF093630) 60S ribosomal protein L21 [Oryza sativa]
- % Identity: 87.2
- Alignment Length: 164
- Location of Alignment in SEQ ID NO 2365: from 1 to 113

Maximum Length Sequence:
related to:
Clone IDs:
253849
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2366
- Ceres seq_id 1572785

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2367
- Ceres seq_id 1572786
- Location of start within SEQ ID NO 2366: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2367: from 201 to 241 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2368
- Ceres seq_id 1572787
- Location of start within SEQ ID NO 2366: at 325 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2368: from 128 to 168 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2369
- Ceres seq_id 1572788
- Location of start within SEQ ID NO 2366: at 472 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2369: from 79 to 119 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
113400
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2370
- Ceres seq_id 1572796

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2371
- Ceres seq_id 1572797
- Location of start within SEQ ID NO 2370: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1094
- gi No. 4115918
- Description: (AF118222) similar to nascent polypeptide associated complex alpha chain [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 156
- Location of Alignment in SEQ ID NO 2371: from 49 to 204

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2372
- Ceres seq_id 1572798
- Location of start within SEQ ID NO 2370: at 330 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1095

- gi No. 4115918
- Description: (AF118222) similar to nascent polypeptide associated complex alpha chain [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 156
- Location of Alignment in SEQ ID NO 2372: from 1 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2373
- Ceres seq_id 1572799
- Location of start within SEQ ID NO 2370: at 345 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1096
- gi No. 4115918
- Description: (AF118222) similar to nascent polypeptide associated complex alpha chain [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 156
- Location of Alignment in SEQ ID NO 2373: from 1 to 133

Maximum Length Sequence:
related to:
Clone IDs:
114364
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2374
- Ceres seq_id 1572800
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2375
- Ceres seq_id 1572801
- Location of start within SEQ ID NO 2374: at 474 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1097
- gi No. 2832644
- Description: (AL021710) teosinte branched1 - like protein [Arabidopsis thaliana]
- % Identity: 86.2
- Alignment Length: 167
- Location of Alignment in SEQ ID NO 2375: from 6 to 172

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2376
- Ceres seq_id 1572802
- Location of start within SEQ ID NO 2374: at 489 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1098
- gi No. 2832644
- Description: (AL021710) teosinte branched1 - like protein [Arabidopsis thaliana]
- % Identity: 86.2
- Alignment Length: 167
- Location of Alignment in SEQ ID NO 2376: from 1 to 167

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2377
- Ceres seq_id 1572803
- Location of start within SEQ ID NO 2374: at 1061 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1099
- gi No. 2832644
- Description: (AL021710) teosinte branched1 - like protein [Arabidopsis thaliana]
- % Identity: 97.3
- Alignment Length: 221
- Location of Alignment in SEQ ID NO 2377: from 1 to 167

Maximum Length Sequence:
related to:
Clone IDs:
115209

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2378
- Ceres seq_id 1572804

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2379
- Ceres seq_id 1572805
- Location of start within SEQ ID NO 2378: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 2379: from 49 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1100
- gi No. 2996650
- Description: (AC004493) KIAA0324 [Homo sapiens]
- % Identity: 70.6
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 2379: from 62 to 95

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2380
- Ceres seq_id 1572806
- Location of start within SEQ ID NO 2378: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 2380: from 30 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1101
- gi No. 2996650
- Description: (AC004493) KIAA0324 [Homo sapiens]
- % Identity: 70.6
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 2380: from 43 to 76

Maximum Length Sequence:

related to:
Clone IDs:
115451
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2381
- Ceres seq_id 1572807
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2382
- Ceres seq_id 1572808
- Location of start within SEQ ID NO 2381: at 339 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 2382: from 18 to 255 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1102
- gi No. 3882356
- Description: (U92460) 12-oxophytodienoate reductase OPR2 [Arabidopsis thaliana]
- % Identity: 70.6
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 2382: from 2 to 258

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2383
- Ceres seq_id 1572809
- Location of start within SEQ ID NO 2381: at 369 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 2383: from 8 to 245 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1103
- gi No. 3882356
- Description: (U92460) 12-oxophytodienoate reductase OPR2 [Arabidopsis thaliana]
- % Identity: 70.6
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 2383: from 1 to 248

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2384
- Ceres seq_id 1572810
- Location of start within SEQ ID NO 2381: at 381 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FMN oxidoreductase
- Location within SEQ ID NO 2384: from 4 to 241 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1104
- gi No. 3882356
- Description: (U92460) 12-oxophytodienoate reductase OPR2 [Arabidopsis thaliana]
- % Identity: 70.6
- Alignment Length: 262

- Location of Alignment in SEQ ID NO 2384: from 1 to 244

Maximum Length Sequence:
related to:
Clone IDs:
115876
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2385
- Ceres seq_id 1572811
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2386
- Ceres seq_id 1572812
- Location of start within SEQ ID NO 2385: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Jacalin-like lectin domain
- Location within SEQ ID NO 2386: from 42 to 99 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2387
- Ceres seq_id 1572813
- Location of start within SEQ ID NO 2385: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
117537
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2388
- Ceres seq_id 1572814
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2389
- Ceres seq_id 1572815
- Location of start within SEQ ID NO 2388: at 599 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 2389: from 35 to 63 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1105
- gi No. 3309082
- Description: (AF076251) calcineurin B-like protein 1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 2389: from 1 to 96

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2390
- Ceres seq_id 1572816
- Location of start within SEQ ID NO 2388: at 641 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 2390: from 21 to 49 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1106
- gi No. 3309082
- Description: (AF076251) calcineurin B-like protein 1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 2390: from 1 to 82

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2391
- Ceres seq_id 1572817
- Location of start within SEQ ID NO 2388: at 671 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 2391: from 11 to 39 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1107
- gi No. 3309082
- Description: (AF076251) calcineurin B-like protein 1 [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 212
- Location of Alignment in SEQ ID NO 2391: from 1 to 72

Maximum Length Sequence:
related to:
Clone IDs:
124735
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2392
- Ceres seq_id 1572826

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2393
- Ceres seq_id 1572827
- Location of start within SEQ ID NO 2392: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1108
- gi No. 3668175
- Description: (AB006778) vegetative storage protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 265
- Location of Alignment in SEQ ID NO 2393: from 14 to 278

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2394
- Ceres seq_id 1572828
- Location of start within SEQ ID NO 2392: at 41 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1109
- gi No. 3668175
- Description: (AB006778) vegetative storage protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 265
- Location of Alignment in SEQ ID NO 2394: from 1 to 265

Maximum Length Sequence:
related to:
Clone IDs:
12830
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2395
- Ceres seq_id 1572833

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2396
- Ceres seq_id 1572834
- Location of start within SEQ ID NO 2395: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2396: from 39 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1110
- gi No. 3164130
- Description: (D78600) cytochrome P450 monooxygenase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 175
- Location of Alignment in SEQ ID NO 2396: from 1 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2397
- Ceres seq_id 1572835
- Location of start within SEQ ID NO 2395: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2397: from 1 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1111
- gi No. 3164130
- Description: (D78600) cytochrome P450 monooxygenase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 175
- Location of Alignment in SEQ ID NO 2397: from 1 to 103

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2398
- Ceres seq_id 1572836
- Location of start within SEQ ID NO 2395: at 676 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2398: from 1 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1112
- gi No. 3164130
- Description: (D78600) cytochrome P450 monooxygenase [Arabidopsis thaliana]
- % Identity: 97.8
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 2398: from 1 to 92

Maximum Length Sequence:
related to:
Clone IDs:
13796
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2399
- Ceres seq_id 1572837
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2400
- Ceres seq_id 1572838
- Location of start within SEQ ID NO 2399: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 2400: from 29 to 196 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1113
- gi No. 4539456
- Description: (AL049500) osmotin precursor [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 196
- Location of Alignment in SEQ ID NO 2400: from 1 to 196

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2401
- Ceres seq_id 1572839
- Location of start within SEQ ID NO 2399: at 220 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 2401: from 1 to 134 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1114
- gi No. 4539456
- Description: (AL049500) osmotin precursor [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 196
- Location of Alignment in SEQ ID NO 2401: from 1 to 134

Maximum Length Sequence:
related to:
Clone IDs:
14210

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2402
- Ceres seq_id 1572840

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2403
- Ceres seq_id 1572841
- Location of start within SEQ ID NO 2402: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1115
- gi No. 2995990
- Description: (AF053746) dormancy-associated protein [Arabidopsis thaliana] >gi|2995992 (AF053747) dormancy-associated protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 2403: from 3 to 124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2404
- Ceres seq_id 1572842
- Location of start within SEQ ID NO 2402: at 9 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1116
- gi No. 2995990
- Description: (AF053746) dormancy-associated protein [Arabidopsis thaliana] >gi|2995992 (AF053747) dormancy-associated protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 2404: from 1 to 122

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2405
- Ceres seq_id 1572843
- Location of start within SEQ ID NO 2402: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1117
- gi No. 2995990
- Description: (AF053746) dormancy-associated protein [Arabidopsis thaliana] >gi|2995992 (AF053747) dormancy-associated protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 122
- Location of Alignment in SEQ ID NO 2405: from 1 to 76

Maximum Length Sequence:
related to:
Clone IDs:
151041

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2406
- Ceres seq_id 1572844

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2407
- Ceres seq_id 1572845
- Location of start within SEQ ID NO 2406: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 2407: from 30 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1118
- gi No. 322571
- Description: proteinase inhibitor II - Arabidopsis thaliana >gi|16427|emb|CAA48892| (X69139) protease inhibitor II [Arabidopsis thaliana] >gi|4038041 (AC005936) proteinase inhibitor II [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 77
- Location of Alignment in SEQ ID NO 2407: from 1 to 76

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2408
- Ceres seq_id 1572846
- Location of start within SEQ ID NO 2406: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 2408: from 15 to 61 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1119
- gi No. 322571
- Description: proteinase inhibitor II - Arabidopsis thaliana >gi|16427|emb|CAA48892| (X69139) protease inhibitor II [Arabidopsis thaliana] >gi|4038041 (AC005936) proteinase inhibitor II [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 77
- Location of Alignment in SEQ ID NO 2408: from 1 to 61

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2409
- Ceres seq_id 1572847
- Location of start within SEQ ID NO 2406: at 311 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
105932
38410

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2410
- Ceres seq_id 1572852

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2411
- Ceres seq_id 1572853
- Location of start within SEQ ID NO 2410: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 14-3-3 proteins
- Location within SEQ ID NO 2411: from 7 to 245 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1120
- gi No. 1345595
- Description: 14-3-3-LIKE PROTEIN GF14 LAMBDA (14-3-3-LIKE PROTEIN AFT1) >gi|1084332|pir||S53727 14-3-3-like protein (ATF1) - Arabidopsis thaliana >gi|953221 (U02565) 14-3-3-like protein 1 [Arabidopsis thaliana] >gi|1549404 (U68545) GF14 lambda
- % Identity: 99.2
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 2411: from 1 to 248

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2412
- Ceres seq_id 1572854
- Location of start within SEQ ID NO 2410: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 14-3-3 proteins
- Location within SEQ ID NO 2412: from 1 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1121
- gi No. 1345595
- Description: 14-3-3-LIKE PROTEIN GF14 LAMBDA (14-3-3-LIKE PROTEIN AFT1) >gi|1084332|pir||S53727 14-3-3-like protein (ATF1) - Arabidopsis thaliana >gi|953221 (U02565) 14-3-3-like protein 1 [Arabidopsis thaliana] >gi|1549404 (U68545) GF14 lambda
- % Identity: 99.2
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 2412: from 1 to 236

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2413
- Ceres seq_id 1572855
- Location of start within SEQ ID NO 2410: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 14-3-3 proteins
- Location within SEQ ID NO 2413: from 1 to 220 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1122
- gi No. 1345595
- Description: 14-3-3-LIKE PROTEIN GF14 LAMBDA (14-3-3-LIKE PROTEIN AFT1) >gi|1084332|pir||S53727 14-3-3-like protein (ATF1) - Arabidopsis thaliana >gi|953221 (U02565) 14-3-3-like protein 1 [Arabidopsis thaliana] >gi|1549404 (U68545) GF14 lambda
- % Identity: 99.2
- Alignment Length: 248
- Location of Alignment in SEQ ID NO 2413: from 1 to 223

Maximum Length Sequence:
related to:

Clone IDs:
17160
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2414
- Ceres seq_id 1572858
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2415
- Ceres seq_id 1572859
- Location of start within SEQ ID NO 2414: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Leucine Rich Repeat
- Location within SEQ ID NO 2415: from 153 to 176 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1123
- gi No. 2827704
- Description: (AL021684) LRR-like protein [Arabidopsis thaliana]
- % Identity: 97.7
- Alignment Length: 257
- Location of Alignment in SEQ ID NO 2415: from 23 to 279

Maximum Length Sequence:
related to:
Clone IDs:
17437
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2416
- Ceres seq_id 1572860
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2417
- Ceres seq_id 1572861
- Location of start within SEQ ID NO 2416: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1124
- gi No. 3250693
- Description: (AL024486) lectin like protein [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2417: from 21 to 266

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2418
- Ceres seq_id 1572862
- Location of start within SEQ ID NO 2416: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1125
- gi No. 3250693
- Description: (AL024486) lectin like protein [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2418: from 1 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2419
- Ceres seq_id 1572863
- Location of start within SEQ ID NO 2416: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1126
- gi No. 3250693
- Description: (AL024486) lectin like protein [Arabidopsis thaliana]
- % Identity: 98.4
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2419: from 1 to 233

Maximum Length Sequence:
related to:
Clone IDs:
21053
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2420
- Ceres seq_id 1572868
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2421
- Ceres seq_id 1572869
- Location of start within SEQ ID NO 2420: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 2421: from 3 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1127
- gi No. 461729
- Description: 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) >gi|2146744|pir||S65597 probable chaperonin, 10K - Arabidopsis thaliana >gi|166662 (L02843) 10 kDa chaperonin [Arabidopsis thaliana]
- % Identity: 74.2
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 2421: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
25828
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2422
- Ceres seq_id 1572878
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2423
- Ceres seq_id 1572879
- Location of start within SEQ ID NO 2422: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2423: from 63 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1128
- gi No. 123539
- Description: 17.5 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.5-E) >gi|169987 (M11395) small heat shock protein [Glycine max]
- % Identity: 78
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 2423: from 22 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2424
- Ceres seq_id 1572880
- Location of start within SEQ ID NO 2422: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2424: from 42 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1129
- gi No. 123539
- Description: 17.5 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.5-E) >gi|169987 (M11395) small heat shock protein [Glycine max]
- % Identity: 78
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 2424: from 1 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2425
- Ceres seq_id 1572881
- Location of start within SEQ ID NO 2422: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2425: from 40 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1130
- gi No. 123539
- Description: 17.5 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.5-E) >gi|169987 (M11395) small heat shock protein [Glycine max]
- % Identity: 78
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 2425: from 1 to 138

Maximum Length Sequence:
related to:
Clone IDs:
27477
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2426
- Ceres seq_id 1572886

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2427
- Ceres seq_id 1572887
- Location of start within SEQ ID NO 2426: at 184 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2427: from 68 to 348 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1131
- gi No. 1171642
- Description: PROBABLE SERINE/THREONINE-PROTEIN KINASE NAK >gi|481206|pir||S38326 protein kinase - Arabidopsis thaliana >gi|166809 (L07248) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 2427: from 1 to 389

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2428
- Ceres seq_id 1572888
- Location of start within SEQ ID NO 2426: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2428: from 29 to 309 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1132
- gi No. 1171642
- Description: PROBABLE SERINE/THREONINE-PROTEIN KINASE NAK >gi|481206|pir||S38326 protein kinase - Arabidopsis thaliana >gi|166809 (L07248) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 2428: from 1 to 350

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2429
- Ceres seq_id 1572889
- Location of start within SEQ ID NO 2426: at 643 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2429: from 1 to 195 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1133
- gi No. 1171642
- Description: PROBABLE SERINE/THREONINE-PROTEIN KINASE NAK >gi|481206|pir||S38326 protein kinase - Arabidopsis thaliana >gi|166809 (L07248) protein kinase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 389
- Location of Alignment in SEQ ID NO 2429: from 1 to 236

Maximum Length Sequence:
related to:
Clone IDs:
27643
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2430
- Ceres seq_id 1572890
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2431
- Ceres seq_id 1572891
- Location of start within SEQ ID NO 2430: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Photosystem I psaG / psaK
- Location within SEQ ID NO 2431: from 71 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1134
- gi No. 3885511
- Description: (AF084200) similar to PSI-K subunit of photosystem I from barley [Medicago sativa]
- % Identity: 80.3
- Alignment Length: 128
- Location of Alignment in SEQ ID NO 2431: from 22 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2432
- Ceres seq_id 1572892
- Location of start within SEQ ID NO 2430: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Photosystem I psaG / psaK
- Location within SEQ ID NO 2432: from 50 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1135
- gi No. 3885511
- Description: (AF084200) similar to PSI-K subunit of photosystem I from barley [Medicago sativa]
- % Identity: 80.3
- Alignment Length: 128
- Location of Alignment in SEQ ID NO 2432: from 1 to 127

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2433
- Ceres seq_id 1572893
- Location of start within SEQ ID NO 2430: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Photosystem I psaG / psaK
- Location within SEQ ID NO 2433: from 46 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1136
- gi No. 3885511
- Description: (AF084200) similar to PSI-K subunit of photosystem I from barley [Medicago sativa]
- % Identity: 80.3
- Alignment Length: 128
- Location of Alignment in SEQ ID NO 2433: from 1 to 123

Maximum Length Sequence:
related to:
Clone IDs:
30238
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2434
- Ceres seq_id 1572894

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2435
- Ceres seq_id 1572895
- Location of start within SEQ ID NO 2434: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione peroxidases
- Location within SEQ ID NO 2435: from 8 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1137
- gi No. 2388887
- Description: (Y14763) glutathione peroxidase [Lycopersicon esculentum]
- % Identity: 75.5
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 2435: from 69 to 166

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2436
- Ceres seq_id 1572896
- Location of start within SEQ ID NO 2434: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione peroxidases
- Location within SEQ ID NO 2436: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1138
- gi No. 2388887
- Description: (Y14763) glutathione peroxidase [Lycopersicon esculentum]
- % Identity: 75.5
- Alignment Length: 98
- Location of Alignment in SEQ ID NO 2436: from 27 to 124

Maximum Length Sequence:
related to:
Clone IDs:
31014

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2437
- Ceres seq_id 1572897

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2438
- Ceres seq_id 1572898
- Location of start within SEQ ID NO 2437: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1139
- gi No. 4884033
- Description: (AJ133753) peptide methionine sulfoxide reductase [Arabidopsis thaliana] >gi|4884035|emb|CAB43187.1| (AJ133754) peptide methionine sulfoxide reductase [Arabidopsis thaliana]
- % Identity: 97.5

- Alignment Length: 204
- Location of Alignment in SEQ ID NO 2438: from 30 to 231

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2439
- Ceres seq_id 1572899
- Location of start within SEQ ID NO 2437: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1140
- gi No. 4884033
- Description: (AJ133753) peptide methionine sulfoxide reductase [Arabidopsis thaliana] >gi|4884035|emb|CAB43187.1| (AJ133754) peptide methionine sulfoxide reductase [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 204
- Location of Alignment in SEQ ID NO 2439: from 1 to 202

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2440
- Ceres seq_id 1572900
- Location of start within SEQ ID NO 2437: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1141
- gi No. 4884033
- Description: (AJ133753) peptide methionine sulfoxide reductase [Arabidopsis thaliana] >gi|4884035|emb|CAB43187.1| (AJ133754) peptide methionine sulfoxide reductase [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 204
- Location of Alignment in SEQ ID NO 2440: from 1 to 186

Maximum Length Sequence:
related to:
Clone IDs:
32602
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2441
- Ceres seq_id 1572905
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2442
- Ceres seq_id 1572906
- Location of start within SEQ ID NO 2441: at 408 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2442: from 84 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1142
- gi No. 4099914
- Description: (U91857) ethylene-responsive element binding protein homolog [Stylosanthes hamata]
- % Identity: 70.2
- Alignment Length: 57

- Location of Alignment in SEQ ID NO 2442: from 87 to 143

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2443
- Ceres seq_id 1572907
- Location of start within SEQ ID NO 2441: at 600 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP2 domain
- Location within SEQ ID NO 2443: from 20 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1143
- gi No. 4099914
- Description: (U91857) ethylene-responsive element binding protein homolog [Stylosanthes hamata]
- % Identity: 70.2
- Alignment Length: 57
- Location of Alignment in SEQ ID NO 2443: from 23 to 79

Maximum Length Sequence:
related to:
Clone IDs:
32789
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2444
- Ceres seq_id 1572912

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2445
- Ceres seq_id 1572913
- Location of start within SEQ ID NO 2444: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1144
- gi No. 1076389
- Description: protein phosphatase 2A pDF1 - Arabidopsis thaliana >gi|683502|emb|CAA57528| (X82002) protein phosphatase 2A 65 kDa regulatory subunit [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 2445: from 1 to 376

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2446
- Ceres seq_id 1572914
- Location of start within SEQ ID NO 2444: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1145
- gi No. 1076389
- Description: protein phosphatase 2A pDF1 - Arabidopsis thaliana >gi|683502|emb|CAA57528| (X82002) protein phosphatase 2A 65 kDa regulatory subunit [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 2446: from 1 to 328

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2447
- Ceres seq_id 1572915
- Location of start within SEQ ID NO 2444: at 397 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1146
- gi No. 1076389
- Description: protein phosphatase 2A pDF1 - Arabidopsis thaliana >gi|683502|emb|CAA57528| (X82002) protein phosphatase 2A 65 kDa regulatory subunit [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 376
- Location of Alignment in SEQ ID NO 2447: from 1 to 244

Maximum Length Sequence:
related to:
Clone IDs:
34405
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2448
- Ceres seq_id 1572916
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2449
- Ceres seq_id 1572917
- Location of start within SEQ ID NO 2448: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2450
- Ceres seq_id 1572918
- Location of start within SEQ ID NO 2448: at 233 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2450: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1147
- gi No. 1169278
- Description: DEHYDRIN ERD14 >gi|556474|dbj|BAA04569| (D17715) ERD14 protein [Arabidopsis thaliana]
- % Identity: 98.9
- Alignment Length: 93
- Location of Alignment in SEQ ID NO 2450: from 1 to 67

Maximum Length Sequence:
related to:
Clone IDs:
36537
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2451
- Ceres seq_id 1572927
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2452
- Ceres seq_id 1572928
- Location of start within SEQ ID NO 2451: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2453
- Ceres seq_id 1572929
- Location of start within SEQ ID NO 2451: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2453: from 1 to 66 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1148
- gi No. 2338712
- Description: (AF013959) metallothionein-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 2453: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
37328

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2454
- Ceres seq_id 1572934

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2455
- Ceres seq_id 1572935
- Location of start within SEQ ID NO 2454: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L35Ae
- Location within SEQ ID NO 2455: from 11 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1149
- gi No. 1709990
- Description: 60S RIBOSOMAL PROTEIN L35A
- % Identity: 100
- Alignment Length: 50
- Location of Alignment in SEQ ID NO 2455: from 1 to 50

Maximum Length Sequence:
related to:
Clone IDs:
38900

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2456
- Ceres seq_id 1572936

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2457

- Ceres seq_id 1572937
- Location of start within SEQ ID NO 2456: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Beta-ketoacyl synthase
- Location within SEQ ID NO 2457: from 88 to 332 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1150
- gi No. 1706749
- Description: 3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE I PRECURSOR (BETA-KETOACYL-ACP SYNTHASE I) (KAS I) >gi|780814 (U24177) 3-ketoacyl-acyl carrier protein synthase I [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 2457: from 26 to 332

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2458
- Ceres seq_id 1572938
- Location of start within SEQ ID NO 2456: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Beta-ketoacyl synthase
- Location within SEQ ID NO 2458: from 63 to 307 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1151
- gi No. 1706749
- Description: 3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE I PRECURSOR (BETA-KETOACYL-ACP SYNTHASE I) (KAS I) >gi|780814 (U24177) 3-ketoacyl-acyl carrier protein synthase I [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 2458: from 1 to 307

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2459
- Ceres seq_id 1572939
- Location of start within SEQ ID NO 2456: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Beta-ketoacyl synthase
- Location within SEQ ID NO 2459: from 1 to 240 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1152
- gi No. 1706749
- Description: 3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE I PRECURSOR (BETA-KETOACYL-ACP SYNTHASE I) (KAS I) >gi|780814 (U24177) 3-ketoacyl-acyl carrier protein synthase I [Arabidopsis thaliana]
- % Identity: 99
- Alignment Length: 307
- Location of Alignment in SEQ ID NO 2459: from 1 to 240

Maximum Length Sequence:
related to:
Clone IDs:

40850
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2460
- Ceres seq_id 1572944

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2461
- Ceres seq_id 1572945
- Location of start within SEQ ID NO 2460: at 794 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DEAD/DEAH box helicase
- Location within SEQ ID NO 2461: from 1 to 53 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1153
- gi No. 3775997
- Description: (AJ010462) RNA helicase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 2461: from 1 to 252

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2462
- Ceres seq_id 1572946
- Location of start within SEQ ID NO 2460: at 953 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helicases conserved C-terminal domain
- Location within SEQ ID NO 2462: from 19 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1154
- gi No. 3775997
- Description: (AJ010462) RNA helicase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 2462: from 1 to 199

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2463
- Ceres seq_id 1572947
- Location of start within SEQ ID NO 2460: at 971 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helicases conserved C-terminal domain
- Location within SEQ ID NO 2463: from 13 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1155
- gi No. 3775997
- Description: (AJ010462) RNA helicase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 289
- Location of Alignment in SEQ ID NO 2463: from 1 to 193

Maximum Length Sequence:
related to:
Clone IDs:

4756
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2464
- Ceres seq_id 1572960
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2465
- Ceres seq_id 1572961
- Location of start within SEQ ID NO 2464: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1156
- gi No. 3367536
- Description: (AC004392) Contains similarity to symbiosis-related like protein F1N20.80 gi|2961343 from A. thaliana BAC gb|AL022140. EST gb|T04695 comes from this gene. [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 128
- Location of Alignment in SEQ ID NO 2465: from 1 to 117

Maximum Length Sequence:
related to:
Clone IDs:
6156
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2466
- Ceres seq_id 1572962
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2467
- Ceres seq_id 1572963
- Location of start within SEQ ID NO 2466: at 439 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L15 amino terminal region
- Location within SEQ ID NO 2467: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1157
- gi No. 132677
- Description: 50S RIBOSOMAL PROTEIN L15, CHLOROPLAST PRECURSOR (CL15) >gi|71241|pir||R5MUL5 ribosomal protein L15 precursor, chloroplast - Arabidopsis thaliana >gi|16497|emb|CAA77593| (Z11508) Plastid ribosomal protein CL15 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 250
- Location of Alignment in SEQ ID NO 2467: from 1 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2468
- Ceres seq_id 1572964
- Location of start within SEQ ID NO 2466: at 475 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L15 amino terminal region
- Location within SEQ ID NO 2468: from 1 to 62 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1158

- gi No. 132677
- Description: 50S RIBOSOMAL PROTEIN L15, CHLOROPLAST PRECURSOR (CL15) >gi|71241|pir||R5MUL5 ribosomal protein L15 precursor, chloroplast - Arabidopsis thaliana >gi|16497|emb|CAA77593| (Z11508) Plastid ribosomal protein CL15 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 250
- Location of Alignment in SEQ ID NO 2468: from 1 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2469
- Ceres seq_id 1572965
- Location of start within SEQ ID NO 2466: at 547 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L15
- Location within SEQ ID NO 2469: from 50 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1159
- gi No. 132677
- Description: 50S RIBOSOMAL PROTEIN L15, CHLOROPLAST PRECURSOR (CL15) >gi|71241|pir||R5MUL5 ribosomal protein L15 precursor, chloroplast - Arabidopsis thaliana >gi|16497|emb|CAA77593| (Z11508) Plastid ribosomal protein CL15 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 250
- Location of Alignment in SEQ ID NO 2469: from 1 to 127

Maximum Length Sequence:
related to:
Clone IDs:
6553
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2470
- Ceres seq_id 1572966

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2471
- Ceres seq_id 1572967
- Location of start within SEQ ID NO 2470: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1160
- gi No. 1732515
- Description: (U62744) myosin heavy chain-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 2471: from 118 to 143

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2472
- Ceres seq_id 1572968
- Location of start within SEQ ID NO 2470: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1161
- gi No. 1732515
- Description: (U62744) myosin heavy chain-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 2472: from 110 to 135

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2473
- Ceres seq_id 1572969
- Location of start within SEQ ID NO 2470: at 296 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1162
- gi No. 1732515
- Description: (U62744) myosin heavy chain-like protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 2473: from 60 to 85

Maximum Length Sequence:
related to:
Clone IDs:
6800
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2474
- Ceres seq_id 1572970
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2475
- Ceres seq_id 1572971
- Location of start within SEQ ID NO 2474: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1163
- gi No. 598073
- Description: (L36806) GT-1 [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 408
- Location of Alignment in SEQ ID NO 2475: from 24 to 431

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2476
- Ceres seq_id 1572972
- Location of start within SEQ ID NO 2474: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1164
- gi No. 598073
- Description: (L36806) GT-1 [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 408
- Location of Alignment in SEQ ID NO 2476: from 1 to 408

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2477
- Ceres seq_id 1572973
- Location of start within SEQ ID NO 2474: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1165
- gi No. 598073
- Description: (L36806) GT-1 [Arabidopsis thaliana]
- % Identity: 98.8
- Alignment Length: 408
- Location of Alignment in SEQ ID NO 2477: from 1 to 380

Maximum Length Sequence:
related to:
Clone IDs:
6984
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2478
- Ceres seq_id 1572974
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2479
- Ceres seq_id 1572975
- Location of start within SEQ ID NO 2478: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Multicopper oxidase
- Location within SEQ ID NO 2479: from 23 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1166
- gi No. 4454012
- Description: (AL035396) Pollen-specific protein precursor like [Arabidopsis thaliana]
- % Identity: 81.4
- Alignment Length: 513
- Location of Alignment in SEQ ID NO 2479: from 1 to 510

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2480
- Ceres seq_id 1572976
- Location of start within SEQ ID NO 2478: at 450 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Multicopper oxidase
- Location within SEQ ID NO 2480: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1167
- gi No. 4454012
- Description: (AL035396) Pollen-specific protein precursor like [Arabidopsis thaliana]
- % Identity: 81.4
- Alignment Length: 513
- Location of Alignment in SEQ ID NO 2480: from 1 to 422

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2481
- Ceres seq_id 1572977
- Location of start within SEQ ID NO 2478: at 747 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1168
- gi No. 4454012
- Description: (AL035396) Pollen-specific protein precursor like [Arabidopsis thaliana]
- % Identity: 81.4
- Alignment Length: 513
- Location of Alignment in SEQ ID NO 2481: from 1 to 323

Maximum Length Sequence:
related to:
Clone IDs:
16766

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2482
- Ceres seq_id 1573108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2483
- Ceres seq_id 1573109
- Location of start within SEQ ID NO 2482: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helicases conserved C-terminal domain
- Location within SEQ ID NO 2483: from 29 to 109 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1169
- gi No. 3776005
- Description: (AJ010466) RNA helicase [Arabidopsis thaliana]
- % Identity: 98
- Alignment Length: 153
- Location of Alignment in SEQ ID NO 2483: from 1 to 153

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2484
- Ceres seq_id 1573110
- Location of start within SEQ ID NO 2482: at 144 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helicases conserved C-terminal domain
- Location within SEQ ID NO 2484: from 1 to 62 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1170
- gi No. 3776005
- Description: (AJ010466) RNA helicase [Arabidopsis thaliana]
- % Identity: 98
- Alignment Length: 153
- Location of Alignment in SEQ ID NO 2484: from 1 to 106

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2485
- Ceres seq_id 1573151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2486
- Ceres seq_id 1573152
- Location of start within SEQ ID NO 2485: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphatidylinositol 3- and 4-kinases
- Location within SEQ ID NO 2486: from 453 to 610 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2487
- Ceres seq_id 1573154
- Location of start within SEQ ID NO 2485: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphatidylinositol 3- and 4-kinases
- Location within SEQ ID NO 2487: from 444 to 601 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2488
- Ceres seq_id 1573308

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2489
- Ceres seq_id 1573309
- Location of start within SEQ ID NO 2488: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 2489: from 233 to 261 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1171
- gi No. 5281051
- Description: (AL080318) stress-induced protein sti1-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 2489: from 1 to 278

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2490
- Ceres seq_id 1573311
- Location of start within SEQ ID NO 2488: at 313 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 2490: from 129 to 157 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1172
- gi No. 5281051
- Description: (AL080318) stress-induced protein sti1-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 2490: from 1 to 174

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2491
- Ceres seq_id 1573315

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2492
- Ceres seq_id 1573316
- Location of start within SEQ ID NO 2491: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2493
- Ceres seq_id 1573317
- Location of start within SEQ ID NO 2491: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1173
- gi No. 2194135
- Description: (AC002062) No definition line found [Arabidopsis thaliana]
- % Identity: 71.1
- Alignment Length: 45
- Location of Alignment in SEQ ID NO 2493: from 1 to 44

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2494
- Ceres seq_id 1573351

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2495
- Ceres seq_id 1573352
- Location of start within SEQ ID NO 2494: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1174
- gi No. 2191155
- Description: (AF007269) contains weak similarity with bacterial SECA proteins. [Arabidopsis thaliana]
- % Identity: 93.9
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 2495: from 1 to 33

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2496
- Ceres seq_id 1573401
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2497
- Ceres seq_id 1573402
- Location of start within SEQ ID NO 2496: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1175
- gi No. 4467359
- Description: (AJ002685) Phosphatidylinositol 4-kinase [Arabidopsis thaliana]
- % Identity: 97.4
- Alignment Length: 38
- Location of Alignment in SEQ ID NO 2497: from 1 to 37

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2498
- Ceres seq_id 1573446
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2499
- Ceres seq_id 1573447
- Location of start within SEQ ID NO 2498: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2499: from 6 to 58 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2500
- Ceres seq_id 1573449
- Location of start within SEQ ID NO 2498: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2501
- Ceres seq_id 1573532
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2502
- Ceres seq_id 1573533
- Location of start within SEQ ID NO 2501: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1176
- gi No. 5724774

- Description: (AF160183) contains similarity to retrotransposons; may be a pseudogene [Arabidopsis thaliana]
- % Identity: 78.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2502: from 60 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2503
- Ceres seq_id 1573535
- Location of start within SEQ ID NO 2501: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1177
- gi No. 5724774
- Description: (AF160183) contains similarity to retrotransposons; may be a pseudogene [Arabidopsis thaliana]
- % Identity: 78.4
- Alignment Length: 163
- Location of Alignment in SEQ ID NO 2503: from 19 to 180

Maximum Length Sequence:
related to:
Clone IDs:
108142
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2504
- Ceres seq_id 1573558
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2505
- Ceres seq_id 1573559
- Location of start within SEQ ID NO 2504: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1178
- gi No. 2134207
- Description: protamine II-1 - painted turtle
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 2505: from 93 to 103

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2506
- Ceres seq_id 1573560
- Location of start within SEQ ID NO 2504: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
20896
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2507
- Ceres seq_id 1573561
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2508
- Ceres seq_id 1573562
- Location of start within SEQ ID NO 2507: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2508: from 168 to 296 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1179
- gi No. 2677828
- Description: (U93166) cysteine protease [Prunus armeniaca]
- % Identity: 74.8
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 2508: from 33 to 294

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2509
- Ceres seq_id 1573563
- Location of start within SEQ ID NO 2507: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2509: from 141 to 269 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1180
- gi No. 2677828
- Description: (U93166) cysteine protease [Prunus armeniaca]
- % Identity: 74.8
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 2509: from 6 to 267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2510
- Ceres seq_id 1573564
- Location of start within SEQ ID NO 2507: at 181 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2510: from 108 to 236 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1181
- gi No. 2677828
- Description: (U93166) cysteine protease [Prunus armeniaca]
- % Identity: 74.8
- Alignment Length: 262
- Location of Alignment in SEQ ID NO 2510: from 1 to 234

Maximum Length Sequence:
related to:
Clone IDs:
14528
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2511
- Ceres seq_id 1573565

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2512
- Ceres seq_id 1573566
- Location of start within SEQ ID NO 2511: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1182
- gi No. 3090403
- Description: (AJ005671) cytochrome b6f complex subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 57
- Location of Alignment in SEQ ID NO 2512: from 30 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2513
- Ceres seq_id 1573567
- Location of start within SEQ ID NO 2511: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1183
- gi No. 3090403
- Description: (AJ005671) cytochrome b6f complex subunit [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 57
- Location of Alignment in SEQ ID NO 2513: from 1 to 57

Maximum Length Sequence:
related to:
Clone IDs:
29669

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2514
- Ceres seq_id 1573568

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2515
- Ceres seq_id 1573569
- Location of start within SEQ ID NO 2514: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2515: from 41 to 384 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1184
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 395
- Location of Alignment in SEQ ID NO 2515: from 28 to 422

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2516
- Ceres seq_id 1573570

- Location of start within SEQ ID NO 2514: at 77 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2516: from 16 to 359 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1185
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 395
- Location of Alignment in SEQ ID NO 2516: from 3 to 397

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2517
- Ceres seq_id 1573571
- Location of start within SEQ ID NO 2514: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2517: from 1 to 336 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1186
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 395
- Location of Alignment in SEQ ID NO 2517: from 1 to 374

Maximum Length Sequence:
related to:
Clone IDs:
8306
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2518
- Ceres seq_id 1573572
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2519
- Ceres seq_id 1573573
- Location of start within SEQ ID NO 2518: at 395 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2519: from 10 to 255 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1187
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100

- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2519: from 10 to 293

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2520
- Ceres seq_id 1573574
- Location of start within SEQ ID NO 2518: at 401 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2520: from 8 to 253 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1188
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2520: from 8 to 291

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2521
- Ceres seq_id 1573575
- Location of start within SEQ ID NO 2518: at 629 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-V
- Location within SEQ ID NO 2521: from 1 to 177 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1189
- gi No. 3288821
- Description: (AF063901) alanine:glyoxylate aminotransferase; transaminase [Arabidopsis thaliana] >gi|4733989|gb|AAD28669.1|AC007209_5 (AC007209) alanine-glyoxylate aminotransferase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2521: from 1 to 215

Maximum Length Sequence:
related to:
Clone IDs:
390

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2522
- Ceres seq_id 1573580

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2523
- Ceres seq_id 1573581
- Location of start within SEQ ID NO 2522: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2524

- Ceres seq_id 1573582
- Location of start within SEQ ID NO 2522: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1190
- gi No. 2765837
- Description: (Z96936) NAP16kDa protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 118
- Location of Alignment in SEQ ID NO 2524: from 1 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2525
- Ceres seq_id 1573583
- Location of start within SEQ ID NO 2522: at 305 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1191
- gi No. 2765837
- Description: (Z96936) NAP16kDa protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 2525: from 39 to 65

Maximum Length Sequence:
related to:
Clone IDs:
25149

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2526
- Ceres seq_id 1573584

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2527
- Ceres seq_id 1573585
- Location of start within SEQ ID NO 2526: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 2527: from 35 to 201 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1192
- gi No. 4678309
- Description: (AL049655) small GTP-binding protein-like (fragment) [Arabidopsis thaliana]
- % Identity: 94.8
- Alignment Length: 194
- Location of Alignment in SEQ ID NO 2527: from 3 to 196

Maximum Length Sequence:
related to:
Clone IDs:
41420

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2528
- Ceres seq_id 1573586

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2529
- Ceres seq_id 1573587
- Location of start within SEQ ID NO 2528: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1193
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 2529: from 22 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2530
- Ceres seq_id 1573588
- Location of start within SEQ ID NO 2528: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1194
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 2530: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2531
- Ceres seq_id 1573589
- Location of start within SEQ ID NO 2528: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1195
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 2531: from 1 to 92

Maximum Length Sequence:
related to:
Clone IDs:
33027
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2532
- Ceres seq_id 1573606
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2533
- Ceres seq_id 1573607
- Location of start within SEQ ID NO 2532: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 2533: from 32 to 131 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2534
- Ceres seq_id 1573608
- Location of start within SEQ ID NO 2532: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 2534: from 27 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2535
- Ceres seq_id 1573609
- Location of start within SEQ ID NO 2532: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 2535: from 15 to 114 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
98440
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2536
- Ceres seq_id 1573624
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2537
- Ceres seq_id 1573625
- Location of start within SEQ ID NO 2536: at 704 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1196
- gi No. 2244759
- Description: (Z97335) selenium-binding protein like [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2537: from 1 to 275

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2538
- Ceres seq_id 1573626
- Location of start within SEQ ID NO 2536: at 821 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1197
- gi No. 2244759
- Description: (Z97335) selenium-binding protein like [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2538: from 1 to 236

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2539
- Ceres seq_id 1573627
- Location of start within SEQ ID NO 2536: at 932 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1198
- gi No. 2244759
- Description: (Z97335) selenium-binding protein like [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 284
- Location of Alignment in SEQ ID NO 2539: from 1 to 199

Maximum Length Sequence:
related to:
Clone IDs:
28007
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2540
- Ceres seq_id 1573628

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2541
- Ceres seq_id 1573629
- Location of start within SEQ ID NO 2540: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2541: from 57 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2542
- Ceres seq_id 1573630
- Location of start within SEQ ID NO 2540: at 46 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2542: from 42 to 111 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2543
- Ceres seq_id 1573631
- Location of start within SEQ ID NO 2540: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2543: from 19 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
40658
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2544
- Ceres seq_id 1573632
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2545
- Ceres seq_id 1573633
- Location of start within SEQ ID NO 2544: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2545: from 58 to 127 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2546
- Ceres seq_id 1573634
- Location of start within SEQ ID NO 2544: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2546: from 42 to 111 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2547
- Ceres seq_id 1573635
- Location of start within SEQ ID NO 2544: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2547: from 19 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34365
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2548
- Ceres seq_id 1573636
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2549
- Ceres seq_id 1573637
- Location of start within SEQ ID NO 2548: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- D-isomer specific 2-hydroxyacid dehydrogenases
- Location within SEQ ID NO 2549: from 123 to 237 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1199
- gi No. 2055273
- Description: (D85339) hydroxypyruvate reductase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 333
- Location of Alignment in SEQ ID NO 2549: from 28 to 360

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2550
- Ceres seq_id 1573638
- Location of start within SEQ ID NO 2548: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- D-isomer specific 2-hydroxyacid dehydrogenases
- Location within SEQ ID NO 2550: from 96 to 210 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1200
- gi No. 2055273
- Description: (D85339) hydroxypyruvate reductase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 333
- Location of Alignment in SEQ ID NO 2550: from 1 to 333

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2551
- Ceres seq_id 1573639
- Location of start within SEQ ID NO 2548: at 152 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- D-isomer specific 2-hydroxyacid dehydrogenases
- Location within SEQ ID NO 2551: from 73 to 187 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1201
- gi No. 2055273
- Description: (D85339) hydroxypyruvate reductase [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 333
- Location of Alignment in SEQ ID NO 2551: from 1 to 310

Maximum Length Sequence:
related to:
Clone IDs:
101930
34351

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2552
- Ceres seq_id 1573640

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2553
- Ceres seq_id 1573641
- Location of start within SEQ ID NO 2552: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- glyceraldehyde 3-phosphate dehydrogenases
- Location within SEQ ID NO 2553: from 15 to 314 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1202
- gi No. 120667
- Description: GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC >gi|81622|pir||JQ1287 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12), cytosolic - Arabidopsis thaliana >gi|166706 (M64116) cystolic glyceraldehyde-3-phosphate
- % Identity: 97.3
- Alignment Length: 338
- Location of Alignment in SEQ ID NO 2553: from 1 to 338

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2554
- Ceres seq_id 1573642
- Location of start within SEQ ID NO 2552: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- glyceraldehyde 3-phosphate dehydrogenases
- Location within SEQ ID NO 2554: from 1 to 270 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1203
- gi No. 120667
- Description: GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC >gi|81622|pir||JQ1287 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12), cytosolic - Arabidopsis thaliana >gi|166706 (M64116) cystolic glyceraldehyde-3-phosphate
- % Identity: 97.3
- Alignment Length: 338
- Location of Alignment in SEQ ID NO 2554: from 1 to 294

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2555
- Ceres seq_id 1573643
- Location of start within SEQ ID NO 2552: at 228 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- glyceraldehyde 3-phosphate dehydrogenases
- Location within SEQ ID NO 2555: from 1 to 267 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1204
- gi No. 120667
- Description: GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC >gi|81622|pir||JQ1287 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12), cytosolic - Arabidopsis thaliana >gi|166706 (M64116) cystolic glyceraldehyde-3-phosphate
- % Identity: 97.3
- Alignment Length: 338

- Location of Alignment in SEQ ID NO 2555: from 1 to 291

Maximum Length Sequence:
related to:
Clone IDs:
255813
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2556
- Ceres seq_id 1573644
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2557
- Ceres seq_id 1573645
- Location of start within SEQ ID NO 2556: at 41 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2557: from 23 to 74 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
206712
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2558
- Ceres seq_id 1573646
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2559
- Ceres seq_id 1573647
- Location of start within SEQ ID NO 2558: at 44 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2559: from 23 to 75 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34995
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2560
- Ceres seq_id 1573648
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2561
- Ceres seq_id 1573649
- Location of start within SEQ ID NO 2560: at 286 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2561: from 89 to 156 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1205
- gi No. 2347192

- Description: (AC002338) spliceosomal protein U2B isolog [Arabidopsis thaliana]
- % Identity: 98.6
- Alignment Length: 218
- Location of Alignment in SEQ ID NO 2561: from 1 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2562
- Ceres seq_id 1573650
- Location of start within SEQ ID NO 2560: at 316 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2562: from 79 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1206
- gi No. 2347192
- Description: (AC002338) spliceosomal protein U2B isolog [Arabidopsis thaliana]
- % Identity: 98.6
- Alignment Length: 218
- Location of Alignment in SEQ ID NO 2562: from 1 to 151

Maximum Length Sequence:
related to:
Clone IDs:
27247
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2563
- Ceres seq_id 1573651

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2564
- Ceres seq_id 1573652
- Location of start within SEQ ID NO 2563: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallo-beta-lactamase superfamily
- Location within SEQ ID NO 2564: from 58 to 228 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1207
- gi No. 1644427
- Description: (U74610) glyoxalase II [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 2564: from 35 to 290

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2565
- Ceres seq_id 1573653
- Location of start within SEQ ID NO 2563: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallo-beta-lactamase superfamily
- Location within SEQ ID NO 2565: from 24 to 194 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1208
- gi No. 1644427
- Description: (U74610) glyoxalase II [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 2565: from 1 to 256

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2566
- Ceres seq_id 1573654
- Location of start within SEQ ID NO 2563: at 308 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallo-beta-lactamase superfamily
- Location within SEQ ID NO 2566: from 1 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1209
- gi No. 1644427
- Description: (U74610) glyoxalase II [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 256
- Location of Alignment in SEQ ID NO 2566: from 1 to 188

Maximum Length Sequence:
related to:
Clone IDs:
30047
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2567
- Ceres seq_id 1573655

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2568
- Ceres seq_id 1573656
- Location of start within SEQ ID NO 2567: at 63 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2568: from 54 to 305 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2569
- Ceres seq_id 1573657
- Location of start within SEQ ID NO 2567: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2569: from 47 to 298 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2570
- Ceres seq_id 1573658
- Location of start within SEQ ID NO 2567: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2570: from 39 to 290 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
37310
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2571
- Ceres seq_id 1573659
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2572
- Ceres seq_id 1573660
- Location of start within SEQ ID NO 2571: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2573
- Ceres seq_id 1573661
- Location of start within SEQ ID NO 2571: at 778 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2573: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2574
- Ceres seq_id 1573662
- Location of start within SEQ ID NO 2571: at 844 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2574: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
39743
38675
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2575
- Ceres seq_id 1573663
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2576
- Ceres seq_id 1573664
- Location of start within SEQ ID NO 2575: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2576: from 54 to 321 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2577
- Ceres seq_id 1573665
- Location of start within SEQ ID NO 2575: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2577: from 47 to 314 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2578
- Ceres seq_id 1573666
- Location of start within SEQ ID NO 2575: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2578: from 39 to 306 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34830
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2579
- Ceres seq_id 1573682

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2580
- Ceres seq_id 1573683
- Location of start within SEQ ID NO 2579: at 155 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1210
- gi No. 4836873
- Description: (AC007260) Highly similar to rice zinc finger protein [Arabidopsis thaliana]
- % Identity: 97.9
- Alignment Length: 96
- Location of Alignment in SEQ ID NO 2580: from 1 to 96

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2581
- Ceres seq_id 1573684
- Location of start within SEQ ID NO 2579: at 532 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1211
- gi No. 4836873
- Description: (AC007260) Highly similar to rice zinc finger protein [Arabidopsis thaliana]
- % Identity: 90.8
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 2581: from 1 to 87

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2582
- Ceres seq_id 1573685
- Location of start within SEQ ID NO 2579: at 577 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1212
- gi No. 4836873
- Description: (AC007260) Highly similar to rice zinc finger protein [Arabidopsis thaliana]
- % Identity: 90.8
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 2582: from 1 to 72

Maximum Length Sequence:
related to:
Clone IDs:
34888

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2583
- Ceres seq_id 1573701

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2584
- Ceres seq_id 1573702
- Location of start within SEQ ID NO 2583: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2584: from 43 to 127 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2585
- Ceres seq_id 1573703
- Location of start within SEQ ID NO 2583: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2585: from 34 to 118 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2586
- Ceres seq_id 1573704
- Location of start within SEQ ID NO 2583: at 400 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
268199
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2587
- Ceres seq_id 1573705
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2588
- Ceres seq_id 1573706
- Location of start within SEQ ID NO 2587: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Asparaginase
- Location within SEQ ID NO 2588: from 20 to 182 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2589
- Ceres seq_id 1573707
- Location of start within SEQ ID NO 2587: at 55 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Asparaginase
- Location within SEQ ID NO 2589: from 2 to 164 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2590
- Ceres seq_id 1573708
- Location of start within SEQ ID NO 2587: at 292 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Asparaginase
- Location within SEQ ID NO 2590: from 1 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
115058
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2591
- Ceres seq_id 1573713
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2592
- Ceres seq_id 1573714
- Location of start within SEQ ID NO 2591: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 2592: from 1 to 181 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
27491
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2593
- Ceres seq_id 1573727
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2594
- Ceres seq_id 1573728
- Location of start within SEQ ID NO 2593: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2595
- Ceres seq_id 1573729
- Location of start within SEQ ID NO 2593: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1213
- gi No. 5730139
- Description: (AJ243705) ferredoxin-NADP+ reductase [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 112
- Location of Alignment in SEQ ID NO 2595: from 1 to 112

Maximum Length Sequence:
related to:
Clone IDs:
149067
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2596
- Ceres seq_id 1573730
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2597
- Ceres seq_id 1573731
- Location of start within SEQ ID NO 2596: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxidoreductase FAD/NAD-binding domain
- Location within SEQ ID NO 2597: from 208 to 302 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1214
- gi No. 5730139
- Description: (AJ243705) ferredoxin-NADP+ reductase [Arabidopsis thaliana]
- % Identity: 98.3
- Alignment Length: 302

- Location of Alignment in SEQ ID NO 2597: from 1 to 302

Maximum Length Sequence:
related to:
Clone IDs:
33021
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2598
- Ceres seq_id 1573732
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2599
- Ceres seq_id 1573733
- Location of start within SEQ ID NO 2598: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1215
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 2599: from 75 to 355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2600
- Ceres seq_id 1573734
- Location of start within SEQ ID NO 2598: at 182 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1216
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 2600: from 61 to 341

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2601
- Ceres seq_id 1573735
- Location of start within SEQ ID NO 2598: at 248 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1217
- gi No. 2821957
- Description: (AB006691) spermidine synthase 2 [Hyoscyamus niger]
- % Identity: 71.3
- Alignment Length: 282
- Location of Alignment in SEQ ID NO 2601: from 39 to 319

Maximum Length Sequence:
related to:
Clone IDs:
124372
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2602
- Ceres seq_id 1573736

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2603
- Ceres seq_id 1573737
- Location of start within SEQ ID NO 2602: at 453 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2603: from 1 to 254 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1218
- gi No. 99699
- Description: glutamate--ammonia ligase (EC 6.3.1.2), cytosolic (clone lambdaAtgsr1) - Arabidopsis thaliana
- % Identity: 98.9
- Alignment Length: 283
- Location of Alignment in SEQ ID NO 2603: from 1 to 266

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2604
- Ceres seq_id 1573738
- Location of start within SEQ ID NO 2602: at 762 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutamine synthetase
- Location within SEQ ID NO 2604: from 1 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1219
- gi No. 99699
- Description: glutamate--ammonia ligase (EC 6.3.1.2), cytosolic (clone lambdaAtgsr1) - Arabidopsis thaliana
- % Identity: 98.9
- Alignment Length: 283
- Location of Alignment in SEQ ID NO 2604: from 1 to 163

Maximum Length Sequence:
related to:
Clone IDs:
37506
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2605
- Ceres seq_id 1573751

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2606
- Ceres seq_id 1573752
- Location of start within SEQ ID NO 2605: at 258 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrogenase E1 component
- Location within SEQ ID NO 2606: from 1 to 296 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1220
- gi No. 5080822
- Description: (AC007258) pyruvate dehydrogenase E1 alpha subunit [Arabidopsis thaliana]
- % Identity: 92.4

- Alignment Length: 354
- Location of Alignment in SEQ ID NO 2606: from 1 to 324

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2607
- Ceres seq_id 1573753
- Location of start within SEQ ID NO 2605: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrogenase E1 component
- Location within SEQ ID NO 2607: from 1 to 293 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1221
- gi No. 5080822
- Description: (AC007258) pyruvate dehydrogenase E1 alpha subunit [Arabidopsis thaliana]
- % Identity: 92.4
- Alignment Length: 354
- Location of Alignment in SEQ ID NO 2607: from 1 to 321

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2608
- Ceres seq_id 1573754
- Location of start within SEQ ID NO 2605: at 276 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrogenase E1 component
- Location within SEQ ID NO 2608: from 1 to 290 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1222
- gi No. 5080822
- Description: (AC007258) pyruvate dehydrogenase E1 alpha subunit [Arabidopsis thaliana]
- % Identity: 92.4
- Alignment Length: 354
- Location of Alignment in SEQ ID NO 2608: from 1 to 318

Maximum Length Sequence:
related to:
Clone IDs:
6281
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2609
- Ceres seq_id 1573759

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2610
- Ceres seq_id 1573760
- Location of start within SEQ ID NO 2609: at 269 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1223
- gi No. 4874274
- Description: (AC007354) Similar to gb|X02844 lipase precursor from Staphylococcus hyicus. ESTs gb|AI239406 and gb|T76725 come from this gene. [Arabidopsis thaliana]

- % Identity: 92.6
- Alignment Length: 230
- Location of Alignment in SEQ ID NO 2610: from 1 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2611
- Ceres seq_id 1573761
- Location of start within SEQ ID NO 2609: at 383 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1224
- gi No. 4874274
- Description: (AC007354) Similar to gb|X02844 lipase precursor from Staphylococcus hyicus. ESTs gb|AI239406 and gb|T76725 come from this gene. [Arabidopsis thaliana]
- % Identity: 92.6
- Alignment Length: 230
- Location of Alignment in SEQ ID NO 2611: from 1 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2612
- Ceres seq_id 1573762
- Location of start within SEQ ID NO 2609: at 407 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1225
- gi No. 4874274
- Description: (AC007354) Similar to gb|X02844 lipase precursor from Staphylococcus hyicus. ESTs gb|AI239406 and gb|T76725 come from this gene. [Arabidopsis thaliana]
- % Identity: 92.6
- Alignment Length: 230
- Location of Alignment in SEQ ID NO 2612: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
207629

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2613
- Ceres seq_id 1573790

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2614
- Ceres seq_id 1573791
- Location of start within SEQ ID NO 2613: at 46 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2614: from 39 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1226
- gi No. 121631
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR >gi|72323|pir||KNNT2S glycine-rich protein 2 - wood tobacco >gi|19743|emb|CAA42622| (X60007) nsGRP-2 [Nicotiana sylvestris]

- % Identity: 70.7
- Alignment Length: 60
- Location of Alignment in SEQ ID NO 2614: from 42 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2615
- Ceres seq_id 1573792
- Location of start within SEQ ID NO 2613: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2615: from 11 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1227
- gi No. 121631
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR >gi|72323|pir||KNNT2S glycine-rich protein 2 - wood tobacco >gi|19743|emb|CAA42622| (X60007) nsGRP-2 [Nicotiana sylvestris]
- % Identity: 70.7
- Alignment Length: 60
- Location of Alignment in SEQ ID NO 2615: from 14 to 71

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2616
- Ceres seq_id 1573793
- Location of start within SEQ ID NO 2613: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
113443
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2617
- Ceres seq_id 1573798

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2618
- Ceres seq_id 1573799
- Location of start within SEQ ID NO 2617: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2618: from 46 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1228
- gi No. 729185
- Description: COLD AND DROUGHT-REGULATED PROTEIN CORA >gi|289123 (L03708) corA gene product [Medicago sativa]
- % Identity: 78.4
- Alignment Length: 38
- Location of Alignment in SEQ ID NO 2618: from 56 to 92

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2619

- Ceres seq_id 1573800
- Location of start within SEQ ID NO 2617: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2619: from 4 to 61 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1229
- gi No. 729185
- Description: COLD AND DROUGHT-REGULATED PROTEIN CORA >gi|289123 (L03708) corA gene product [Medicago sativa]
- % Identity: 78.4
- Alignment Length: 38
- Location of Alignment in SEQ ID NO 2619: from 14 to 50

Maximum Length Sequence:
related to:
Clone IDs:
17376
34326
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2620
- Ceres seq_id 1573801
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2621
- Ceres seq_id 1573802
- Location of start within SEQ ID NO 2620: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2621: from 29 to 139 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1230
- gi No. 166374
- Description: (M74190) environmental stress-induced protein [Medicago sativa]
- % Identity: 70.2
- Alignment Length: 47
- Location of Alignment in SEQ ID NO 2621: from 71 to 117

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2622
- Ceres seq_id 1573803
- Location of start within SEQ ID NO 2620: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ELM2 domain
- Location within SEQ ID NO 2622: from 29 to 109 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1231
- gi No. 259447
- Description: (S47409) glycine-rich protein, atGRP {clone atGRP-3} [Arabidopsis thaliana, C24, Peptide, 145 aa] [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 35

- Location of Alignment in SEQ ID NO 2622: from 1 to 35

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2623
- Ceres seq_id 1573804
- Location of start within SEQ ID NO 2620: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ELM2 domain
- Location within SEQ ID NO 2623: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
102088
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2624
- Ceres seq_id 1573809
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2625
- Ceres seq_id 1573810
- Location of start within SEQ ID NO 2624: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2625: from 63 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1232
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 70.7
- Alignment Length: 77
- Location of Alignment in SEQ ID NO 2625: from 59 to 130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2626
- Ceres seq_id 1573811
- Location of start within SEQ ID NO 2624: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2626: from 18 to 85 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1233
- gi No. 544424
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 7 >gi|419755|pir||S30147 glycine-rich protein (clone AtGRP7) - Arabidopsis thaliana >gi|16301|emb|CAA78711| (Z14987) glycine rich protein [Arabidopsis thaliana] >gi|166837 (L00648) RNA-binding thaliana]
- % Identity: 70.7
- Alignment Length: 77

- Location of Alignment in SEQ ID NO 2626: from 14 to 85

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2627
- Ceres seq_id 1573812
- Location of start within SEQ ID NO 2624: at 191 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
3929
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2628
- Ceres seq_id 1573813
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2629
- Ceres seq_id 1573814
- Location of start within SEQ ID NO 2628: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2629: from 14 to 89 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1234
- gi No. 585218
- Description: GLYCINE-RICH PROTEIN DC7.1 PRECURSOR >gi|486810|pir||S35715 glycine-rich protein (clone DC 7.1), embryonic - carrot >gi|18349|emb|CAA33736| (X15706) glycine rich protein (AA 1 - 96) [Daucus carota]
- % Identity: 71.9
- Alignment Length: 32
- Location of Alignment in SEQ ID NO 2629: from 50 to 81

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2630
- Ceres seq_id 1573815
- Location of start within SEQ ID NO 2628: at 195 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2631
- Ceres seq_id 1573816
- Location of start within SEQ ID NO 2628: at 228 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
100976

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2632
- Ceres seq_id 1573817

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2633
- Ceres seq_id 1573818
- Location of start within SEQ ID NO 2632: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2633: from 3 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1235
- gi No. 585218
- Description: GLYCINE-RICH PROTEIN DC7.1 PRECURSOR >gi|486810|pir||S35715 glycine-rich protein (clone DC 7.1), embryonic - carrot >gi|18349|emb|CAA33736| (X15706) glycine rich protein (AA 1 - 96) [Daucus carota]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 2633: from 54 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2634
- Ceres seq_id 1573819
- Location of start within SEQ ID NO 2632: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2634: from 1 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1236
- gi No. 585218
- Description: GLYCINE-RICH PROTEIN DC7.1 PRECURSOR >gi|486810|pir||S35715 glycine-rich protein (clone DC 7.1), embryonic - carrot >gi|18349|emb|CAA33736| (X15706) glycine rich protein (AA 1 - 96) [Daucus carota]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 2634: from 51 to 87

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2635
- Ceres seq_id 1573820
- Location of start within SEQ ID NO 2632: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ELM2 domain
- Location within SEQ ID NO 2635: from 1 to 64 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
109026

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2636
- Ceres seq_id 1573821

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2637
- Ceres seq_id 1573822
- Location of start within SEQ ID NO 2636: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2637: from 1 to 106 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1237
- gi No. 4006934
- Description: (AJ012571) glutathione transferase [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 40
- Location of Alignment in SEQ ID NO 2637: from 95 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2638
- Ceres seq_id 1573823
- Location of start within SEQ ID NO 2636: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ELM2 domain
- Location within SEQ ID NO 2638: from 1 to 64 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
106266

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2639
- Ceres seq_id 1573832

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2640
- Ceres seq_id 1573833
- Location of start within SEQ ID NO 2639: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 2640: from 88 to 276 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1238
- gi No. 400976
- Description: RAS-RELATED PROTEIN RHA1 >gi|478671|pir||S23727 GTP-binding protein RHA1 - Arabidopsis thaliana >gi|16484|emb|CAA41863| (X59152) RHA1 [Arabidopsis thaliana] >gi|397594|emb|CAA80534| (Z22958) GTP-binding protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2640: from 77 to 276

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2641
- Ceres seq_id 1573834
- Location of start within SEQ ID NO 2639: at 230 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ras family
- Location within SEQ ID NO 2641: from 12 to 200 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1239
- gi No. 400976
- Description: RAS-RELATED PROTEIN RHA1 >gi|478671|pir||S23727 GTP-binding protein RHA1 - Arabidopsis thaliana >gi|16484|emb|CAA41863| (X59152) RHA1 [Arabidopsis thaliana] >gi|397594|emb|CAA80534| (Z22958) GTP-binding protein [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2641: from 1 to 200

Maximum Length Sequence:
related to:
Clone IDs:
112971
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2642
- Ceres seq_id 1573857
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2643
- Ceres seq_id 1573858
- Location of start within SEQ ID NO 2642: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 2643: from 2 to 70 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2644
- Ceres seq_id 1573859
- Location of start within SEQ ID NO 2642: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2645
- Ceres seq_id 1573860
- Location of start within SEQ ID NO 2642: at 44 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 2645: from 1 to 56 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
155326
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2646
- Ceres seq_id 1573864
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2647
- Ceres seq_id 1573865
- Location of start within SEQ ID NO 2646: at 190 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2647: from 110 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1240
- gi No. 2108256
- Description: (Y13141) extensin [Bromheadia finlaysoniana]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 2647: from 13 to 23

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2648
- Ceres seq_id 1573866
- Location of start within SEQ ID NO 2646: at 304 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2648: from 72 to 111 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
2955
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2649
- Ceres seq_id 1573867
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2650
- Ceres seq_id 1573868
- Location of start within SEQ ID NO 2649: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 2650: from 2 to 126 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2651
- Ceres seq_id 1573869
- Location of start within SEQ ID NO 2649: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 2651: from 1 to 95 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
6135
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2652
- Ceres seq_id 1573899
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2653
- Ceres seq_id 1573900
- Location of start within SEQ ID NO 2652: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2653: from 5 to 405 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1241
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 2653: from 1 to 449

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2654
- Ceres seq_id 1573901
- Location of start within SEQ ID NO 2652: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2654: from 1 to 357 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1242
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 2654: from 1 to 401

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2655
- Ceres seq_id 1573902
- Location of start within SEQ ID NO 2652: at 385 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 2655: from 1 to 304 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1243
- gi No. 119143
- Description: ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gi|81606|pir||S06724 translation elongation factor eEF-1 alpha chain - Arabidopsis thaliana >gi|295788|emb|CAA34453| (X16430) elongation factor 1-alpha [Arabidopsis thaliana]
- % Identity: 99.8
- Alignment Length: 449
- Location of Alignment in SEQ ID NO 2655: from 1 to 348

Maximum Length Sequence:
related to:
Clone IDs:
35452
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2656
- Ceres seq_id 1573909
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2657
- Ceres seq_id 1573910
- Location of start within SEQ ID NO 2656: at 52 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1244
- gi No. 3355475
- Description: (AC004218) ribosomal protein L23a [Arabidopsis thaliana]
- % Identity: 75
- Alignment Length: 60
- Location of Alignment in SEQ ID NO 2657: from 1 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2658
- Ceres seq_id 1573911
- Location of start within SEQ ID NO 2656: at 305 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L23
- Location within SEQ ID NO 2658: from 1 to 67 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1245
- gi No. 3355475
- Description: (AC004218) ribosomal protein L23a [Arabidopsis thaliana]
- % Identity: 97.2
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 2658: from 1 to 70

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2659

- Ceres seq_id 1573912
- Location of start within SEQ ID NO 2656: at 392 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1246
- gi No. 3355475
- Description: (AC004218) ribosomal protein L23a [Arabidopsis thaliana]
- % Identity: 97.2
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 2659: from 1 to 41

Maximum Length Sequence:
related to:
Clone IDs:
12507
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2660
- Ceres seq_id 1573920

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2661
- Ceres seq_id 1573921
- Location of start within SEQ ID NO 2660: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2661: from 28 to 80 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1247
- gi No. 2497886
- Description: METALLOTHIONEIN-LIKE PROTEIN 2B (MT-2B) >gi|1361999|pir||S57862 metallothionein 2b - Arabidopsis thaliana >gi|1086463 (U11256) metallothionein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 53
- Location of Alignment in SEQ ID NO 2661: from 28 to 80

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2662
- Ceres seq_id 1573922
- Location of start within SEQ ID NO 2660: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2662: from 1 to 53 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1248
- gi No. 2497886
- Description: METALLOTHIONEIN-LIKE PROTEIN 2B (MT-2B) >gi|1361999|pir||S57862 metallothionein 2b - Arabidopsis thaliana >gi|1086463 (U11256) metallothionein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 53
- Location of Alignment in SEQ ID NO 2662: from 1 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2663
- Ceres seq_id 1573923
- Location of start within SEQ ID NO 2660: at 358 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
30015

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2664
- Ceres seq_id 1573941

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2665
- Ceres seq_id 1573942
- Location of start within SEQ ID NO 2664: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2665: from 46 to 192 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2666
- Ceres seq_id 1573943
- Location of start within SEQ ID NO 2664: at 10 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2666: from 43 to 189 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2667
- Ceres seq_id 1573944
- Location of start within SEQ ID NO 2664: at 175 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Lipase/Acylhydrolase with GDSL-like motif
- Location within SEQ ID NO 2667: from 1 to 134 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
115663

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2668
- Ceres seq_id 1573949

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2669

- Ceres seq_id 1573950
- Location of start within SEQ ID NO 2668: at 162 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2669: from 111 to 167 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2670
- Ceres seq_id 1573951
- Location of start within SEQ ID NO 2668: at 180 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2670: from 105 to 161 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2671
- Ceres seq_id 1573952
- Location of start within SEQ ID NO 2668: at 691 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1249
- gi No. 1666299
- Description: (X92980) RNA-binding protein [Anabaena variabilis]
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 2671: from 77 to 92

Maximum Length Sequence:
related to:
Clone IDs:
32376
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2672
- Ceres seq_id 1573953

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2673
- Ceres seq_id 1573954
- Location of start within SEQ ID NO 2672: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- tRNA synthetases class II (G, H, P and S)
- Location within SEQ ID NO 2673: from 1 to 109 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1250
- gi No. 3334349
- Description: GLYCYL-TRNA SYNTHETASE (GLYCINE--TRNA LIGASE) (GLYRS) >gi|2564215|emb|CAA05162| (AJ002062) glycyl-tRNA synthetase [Arabidopsis thaliana]
- % Identity: 99.3

- Alignment Length: 147
- Location of Alignment in SEQ ID NO 2673: from 1 to 147

Maximum Length Sequence:
related to:
Clone IDs:
157869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2674
- Ceres seq_id 1573955
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2675
- Ceres seq_id 1573956
- Location of start within SEQ ID NO 2674: at 289 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2675: from 74 to 238 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1251
- gi No. 1172872
- Description: CYSTEINE PROTEINASE RD19A PRECURSOR >gi|541856|pir||JN0718 drought-inducible cysteine proteinase (EC 3.4.22.-) RD19A precursor - Arabidopsis thaliana >gi|435618|dbj|BAA02373| (D13042) thiol protease [Arabidopsis
- % Identity: 97.7
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 2675: from 74 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2676
- Ceres seq_id 1573957
- Location of start within SEQ ID NO 2674: at 352 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2676: from 53 to 217 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1252
- gi No. 1172872
- Description: CYSTEINE PROTEINASE RD19A PRECURSOR >gi|541856|pir||JN0718 drought-inducible cysteine proteinase (EC 3.4.22.-) RD19A precursor - Arabidopsis thaliana >gi|435618|dbj|BAA02373| (D13042) thiol protease [Arabidopsis
- % Identity: 97.7
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 2676: from 53 to 225

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2677
- Ceres seq_id 1573958
- Location of start within SEQ ID NO 2674: at 553 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Papain family cysteine protease
- Location within SEQ ID NO 2677: from 1 to 150 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1253
- gi No. 1172872
- Description: CYSTEINE PROTEINASE RD19A PRECURSOR >gi|541856|pir||JN0718 drought-inducible cysteine proteinase (EC 3.4.22.-) RD19A precursor - Arabidopsis thaliana >gi|435618|dbj|BAA02373| (D13042) thiol protease [Arabidopsis
- % Identity: 97.7
- Alignment Length: 173
- Location of Alignment in SEQ ID NO 2677: from 1 to 158

Maximum Length Sequence:
related to:
Clone IDs:
35584
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2678
- Ceres seq_id 1573960
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2679
- Ceres seq_id 1573961
- Location of start within SEQ ID NO 2678: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ADP-ribosylation factor family
- Location within SEQ ID NO 2679: from 2 to 180 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1254
- gi No. 543841
- Description: ADP-RIBOSYLATION FACTOR 1 >gi|322518|pir||S28875 ADP-ribosylation factor 1 - Arabidopsis thaliana >gi|166586 (M95166) ADP-ribosylation factor [Arabidopsis thaliana] >gi|2275195 (AC002337) ADP-ribosylation factor
- % Identity: 100
- Alignment Length: 181
- Location of Alignment in SEQ ID NO 2679: from 1 to 181

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2680
- Ceres seq_id 1573962
- Location of start within SEQ ID NO 2678: at 167 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ADP-ribosylation factor family
- Location within SEQ ID NO 2680: from 1 to 163 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1255
- gi No. 543841
- Description: ADP-RIBOSYLATION FACTOR 1 >gi|322518|pir||S28875 ADP-ribosylation factor 1 - Arabidopsis thaliana >gi|166586 (M95166) ADP-ribosylation factor [Arabidopsis thaliana] >gi|2275195 (AC002337) ADP-ribosylation factor
- % Identity: 100
- Alignment Length: 181
- Location of Alignment in SEQ ID NO 2680: from 1 to 164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2681
- Ceres seq_id 1573963
- Location of start within SEQ ID NO 2678: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ADP-ribosylation factor family
- Location within SEQ ID NO 2681: from 1 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1256
- gi No. 543841
- Description: ADP-RIBOSYLATION FACTOR 1 >gi|322518|pir||S28875 ADP-ribosylation factor 1 - Arabidopsis thaliana >gi|166586 (M95166) ADP-ribosylation factor [Arabidopsis thaliana] >gi|2275195 (AC002337) ADP-ribosylation factor
- % Identity: 100
- Alignment Length: 181
- Location of Alignment in SEQ ID NO 2681: from 1 to 160

Maximum Length Sequence:
related to:
Clone IDs:
28471
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2682
- Ceres seq_id 1573964
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2683
- Ceres seq_id 1573965
- Location of start within SEQ ID NO 2682: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 2683: from 144 to 212 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1257
- gi No. 548653
- Description: 50S RIBOSOMAL PROTEIN L12-A, CHLOROPLAST PRECURSOR (CL12-A) >gi|541895|pir||A53394 ribosomal protein L12.A, chloroplast - Arabidopsis thaliana >gi|468771|emb|CAA48181| (X68046) ribosomal protein L12 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 191
- Location of Alignment in SEQ ID NO 2683: from 22 to 212

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2684
- Ceres seq_id 1573966
- Location of start within SEQ ID NO 2682: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 2684: from 123 to 191 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1258

- gi No. 548653
- Description: 50S RIBOSOMAL PROTEIN L12-A, CHLOROPLAST PRECURSOR (CL12-A) >gi|541895|pir||A53394 ribosomal protein L12.A, chloroplast - Arabidopsis thaliana >gi|468771|emb|CAA48181| (X68046) ribosomal protein L12 [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 191
- Location of Alignment in SEQ ID NO 2684: from 1 to 191

Maximum Length Sequence:
related to:
Clone IDs:
6966
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2685
- Ceres seq_id 1573980

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2686
- Ceres seq_id 1573981
- Location of start within SEQ ID NO 2685: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 2686: from 69 to 122 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1259
- gi No. 3077806
- Description: (AB013076) loricrin [Homo sapiens]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 2686: from 70 to 90

Maximum Length Sequence:
related to:
Clone IDs:
13439
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2687
- Ceres seq_id 1573982

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2688
- Ceres seq_id 1573983
- Location of start within SEQ ID NO 2687: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2688: from 8 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1260
- gi No. 544425
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 100
- Alignment Length: 159
- Location of Alignment in SEQ ID NO 2688: from 1 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2689
- Ceres seq_id 1573984
- Location of start within SEQ ID NO 2687: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2689: from 29 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1261
- gi No. 544425
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 100
- Alignment Length: 159
- Location of Alignment in SEQ ID NO 2689: from 1 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2690
- Ceres seq_id 1573985
- Location of start within SEQ ID NO 2687: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dehydrins
- Location within SEQ ID NO 2690: from 22 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1262
- gi No. 544425
- Description: GLYCINE-RICH RNA-BINDING PROTEIN 8 (CCR1 PROTEIN) >gi|419756|pir||S30148 glycine-rich protein (clone AtGRP8) - Arabidopsis thaliana >gi|16305|emb|CAA78712| (Z14988) glycine rich protein [Arabidopsis thaliana] >gi|166658 (L04171) ORF
- % Identity: 100
- Alignment Length: 159
- Location of Alignment in SEQ ID NO 2690: from 1 to 93

Maximum Length Sequence:
related to:
Clone IDs:
231128
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2691
- Ceres seq_id 1573986

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2692
- Ceres seq_id 1573987
- Location of start within SEQ ID NO 2691: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 2692: from 11 to 69 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1263
- gi No. 543815
- Description: FLORAL HOMEOTIC PROTEIN APETALA3 >gi|282855|pir||A42095 homeotic protein AP3 - Arabidopsis thaliana >gi|166608 (M86357) APETELA3 [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 168
- Location of Alignment in SEQ ID NO 2692: from 11 to 178

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2693
- Ceres seq_id 1573988
- Location of start within SEQ ID NO 2691: at 31 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SRF-type transcription factor (DNA-binding and dimerisation domain)
- Location within SEQ ID NO 2693: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1264
- gi No. 543815
- Description: FLORAL HOMEOTIC PROTEIN APETALA3 >gi|282855|pir||A42095 homeotic protein AP3 - Arabidopsis thaliana >gi|166608 (M86357) APETELA3 [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 168
- Location of Alignment in SEQ ID NO 2693: from 1 to 168

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2694
- Ceres seq_id 1573989
- Location of start within SEQ ID NO 2691: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1265
- gi No. 543815
- Description: FLORAL HOMEOTIC PROTEIN APETALA3 >gi|282855|pir||A42095 homeotic protein AP3 - Arabidopsis thaliana >gi|166608 (M86357) APETELA3 [Arabidopsis thaliana]
- % Identity: 99.4
- Alignment Length: 168
- Location of Alignment in SEQ ID NO 2694: from 1 to 122

Maximum Length Sequence:
related to:
Clone IDs:
31731

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2695
- Ceres seq_id 1573990

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2696
- Ceres seq_id 1573991
- Location of start within SEQ ID NO 2695: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2697
- Ceres seq_id 1573992
- Location of start within SEQ ID NO 2695: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1266
- gi No. 4587564
- Description: (AC006550) Strong similarity to gb|X14017 photosystem I reaction centre subunit II precursor (psaD) from Spinacia oleracea. ESTs gb|R30423, gb|T42998, gb|Z18178, gb|T14133, gb|N65521, gb|T42498, gb|T41918, gb|N38024...
- % Identity: 93.3
- Alignment Length: 208
- Location of Alignment in SEQ ID NO 2697: from 1 to 208

Maximum Length Sequence:
related to:
Clone IDs:
19707
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2698
- Ceres seq_id 1573995
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2699
- Ceres seq_id 1573996
- Location of start within SEQ ID NO 2698: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1267
- gi No. 2464855
- Description: (Z99707) myb-related protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2699: from 1 to 189

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2700
- Ceres seq_id 1573997
- Location of start within SEQ ID NO 2698: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1268
- gi No. 2464855
- Description: (Z99707) myb-related protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2700: from 1 to 166

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2701

- Ceres seq_id 1573998
- Location of start within SEQ ID NO 2698: at 272 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1269
- gi No. 2464855
- Description: (Z99707) myb-related protein [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 2701: from 1 to 99

Maximum Length Sequence:
related to:
Clone IDs:
36801
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2702
- Ceres seq_id 1573999
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2703
- Ceres seq_id 1574000
- Location of start within SEQ ID NO 2702: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1270
- gi No. 4741966
- Description: (AF134133) Lil3 protein [Arabidopsis thaliana]
- % Identity: 77.9
- Alignment Length: 254
- Location of Alignment in SEQ ID NO 2703: from 41 to 286

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2704
- Ceres seq_id 1574001
- Location of start within SEQ ID NO 2702: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1271
- gi No. 4741966
- Description: (AF134133) Lil3 protein [Arabidopsis thaliana]
- % Identity: 77.9
- Alignment Length: 254
- Location of Alignment in SEQ ID NO 2704: from 5 to 250

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2705
- Ceres seq_id 1574002
- Location of start within SEQ ID NO 2702: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1272
- gi No. 4741966
- Description: (AF134133) Lil3 protein [Arabidopsis thaliana]

- % Identity: 77.9
- Alignment Length: 254
- Location of Alignment in SEQ ID NO 2705: from 1 to 246

Maximum Length Sequence:
related to:
Clone IDs:
17426
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2706
- Ceres seq_id 1574006
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2707
- Ceres seq_id 1574007
- Location of start within SEQ ID NO 2706: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1273
- gi No. 3157948
- Description: (AC002131) Contains similarity to thyroidal receptor for N-acetylglucosamine gb|A27284 from Homo sapiens. EST gb|N65493 comes from this gene. [Arabidopsis thaliana]
- % Identity: 93.5
- Alignment Length: 92
- Location of Alignment in SEQ ID NO 2707: from 12 to 101

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2708
- Ceres seq_id 1574008
- Location of start within SEQ ID NO 2706: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1274
- gi No. 3157948
- Description: (AC002131) Contains similarity to thyroidal receptor for N-acetylglucosamine gb|A27284 from Homo sapiens. EST gb|N65493 comes from this gene. [Arabidopsis thaliana]
- % Identity: 93.5
- Alignment Length: 92
- Location of Alignment in SEQ ID NO 2708: from 1 to 90

Maximum Length Sequence:
related to:
Clone IDs:
2132
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2709
- Ceres seq_id 1574009
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2710
- Ceres seq_id 1574010
- Location of start within SEQ ID NO 2709: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 2710: from 87 to 275 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1275
- gi No. 5360589
- Description: (AB022327) nClpP2 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 2710: from 17 to 295

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2711
- Ceres seq_id 1574011
- Location of start within SEQ ID NO 2709: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 2711: from 71 to 259 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1276
- gi No. 5360589
- Description: (AB022327) nClpP2 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 2711: from 1 to 279

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2712
- Ceres seq_id 1574012
- Location of start within SEQ ID NO 2709: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clp protease
- Location within SEQ ID NO 2712: from 5 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1277
- gi No. 5360589
- Description: (AB022327) nClpP2 [Arabidopsis thaliana]
- % Identity: 99.3
- Alignment Length: 279
- Location of Alignment in SEQ ID NO 2712: from 1 to 213

Maximum Length Sequence:
related to:
Clone IDs:
19036
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2713
- Ceres seq_id 1574013

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2714
- Ceres seq_id 1574014
- Location of start within SEQ ID NO 2713: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Enoyl-CoA hydratase/isomerase family
- Location within SEQ ID NO 2714: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2715
- Ceres seq_id 1574015
- Location of start within SEQ ID NO 2713: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
22936

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2716
- Ceres seq_id 1574021

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2717
- Ceres seq_id 1574022
- Location of start within SEQ ID NO 2716: at 157 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2717: from 48 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1278
- gi No. 5007080
- Description: (AF153689) poly(A)-binding protein [Oryza sativa]
- % Identity: 81.8
- Alignment Length: 22
- Location of Alignment in SEQ ID NO 2717: from 280 to 301

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2718
- Ceres seq_id 1574023
- Location of start within SEQ ID NO 2716: at 932 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2718: from 72 to 142 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2719
- Ceres seq_id 1574024
- Location of start within SEQ ID NO 2716: at 938 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2719: from 70 to 140 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
34633
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2720
- Ceres seq_id 1574025
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2721
- Ceres seq_id 1574026
- Location of start within SEQ ID NO 2720: at 52 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2721: from 17 to 160 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1279
- gi No. 1172599
- Description: PROTEASOME COMPONENT C5 (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C5) (TAS-F22/FAFP98) >gi|600387|emb|CAA47753| (X67338) proteosome subunit [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2721: from 1 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2722
- Ceres seq_id 1574027
- Location of start within SEQ ID NO 2720: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2722: from 1 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1280
- gi No. 1172599
- Description: PROTEASOME COMPONENT C5 (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C5) (TAS-F22/FAFP98) >gi|600387|emb|CAA47753| (X67338) proteosome subunit [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2722: from 1 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2723
- Ceres seq_id 1574028
- Location of start within SEQ ID NO 2720: at 316 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2723: from 1 to 72 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1281
- gi No. 1172599

- Description: PROTEASOME COMPONENT C5 (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C5) (TAS-F22/FAFP98) >gi|600387|emb|CAA47753| (X67338) proteosome subunit [Arabidopsis thaliana]
- % Identity: 99.5
- Alignment Length: 200
- Location of Alignment in SEQ ID NO 2723: from 1 to 105

Maximum Length Sequence:
related to:
Clone IDs:
33783
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2724
- Ceres seq_id 1574029

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2725
- Ceres seq_id 1574030
- Location of start within SEQ ID NO 2724: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2725: from 1 to 309 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1282
- gi No. 5042165
- Description: (AL078620) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2725: from 157 to 310

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2726
- Ceres seq_id 1574031
- Location of start within SEQ ID NO 2724: at 5 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2726: from 1 to 308 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1283
- gi No. 5042165
- Description: (AL078620) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2726: from 156 to 309

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2727
- Ceres seq_id 1574032
- Location of start within SEQ ID NO 2724: at 23 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 2727: from 1 to 302 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1284
- gi No. 5042165
- Description: (AL078620) cytochrome P450-like protein [Arabidopsis thaliana]
- % Identity: 70.1
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 2727: from 150 to 303

Maximum Length Sequence:
related to:
Clone IDs:
34078
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2728
- Ceres seq_id 1574038
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2729
- Ceres seq_id 1574039
- Location of start within SEQ ID NO 2728: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type class I peptide chain release factors
- Location within SEQ ID NO 2729: from 47 to 149 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2730
- Ceres seq_id 1574040
- Location of start within SEQ ID NO 2728: at 405 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type class I peptide chain release factors
- Location within SEQ ID NO 2730: from 1 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
35476
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2731
- Ceres seq_id 1574041
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2732
- Ceres seq_id 1574042
- Location of start within SEQ ID NO 2731: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1285
- gi No. 2209332
- Description: (U89272) chloroplast membrane protein ALBINO3 [Arabidopsis thaliana] >gi|3927828 (AC005727) chloroplast membrane protein ALBINO3 [Arabidopsis thaliana]

- % Identity: 99.2
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 2732: from 1 to 130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2733
- Ceres seq_id 1574043
- Location of start within SEQ ID NO 2731: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1286
- gi No. 2209332
- Description: (U89272) chloroplast membrane protein ALBINO3 [Arabidopsis thaliana] >gi|3927828 (AC005727) chloroplast membrane protein ALBINO3 [Arabidopsis thaliana]
- % Identity: 99.2
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 2733: from 1 to 101

Maximum Length Sequence:
related to:
Clone IDs:
109373
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2734
- Ceres seq_id 1574044
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2735
- Ceres seq_id 1574045
- Location of start within SEQ ID NO 2734: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ubiE/COQ5 methyltransferase family
- Location within SEQ ID NO 2735: from 66 to 198 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2736
- Ceres seq_id 1574046
- Location of start within SEQ ID NO 2734: at 149 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ubiE/COQ5 methyltransferase family
- Location within SEQ ID NO 2736: from 38 to 170 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
42713
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2737
- Ceres seq_id 1574047
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2738

- Ceres seq_id 1574048
- Location of start within SEQ ID NO 2737: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2738: from 54 to 321 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1287
- gi No. 3641314
- Description: (AF087413) AJH1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2738: from 1 to 357

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2739
- Ceres seq_id 1574049
- Location of start within SEQ ID NO 2737: at 305 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2739: from 1 to 253 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1288
- gi No. 3641314
- Description: (AF087413) AJH1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2739: from 1 to 289

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2740
- Ceres seq_id 1574050
- Location of start within SEQ ID NO 2737: at 344 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mov34 family
- Location within SEQ ID NO 2740: from 1 to 240 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1289
- gi No. 3641314
- Description: (AF087413) AJH1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 357
- Location of Alignment in SEQ ID NO 2740: from 1 to 276

Maximum Length Sequence:
related to:
Clone IDs:
113419
3797
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2741
- Ceres seq_id 1574059
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2742
- Ceres seq_id 1574060
- Location of start within SEQ ID NO 2741: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 2742: from 109 to 215 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1290
- gi No. 99991
- Description: protein disulfide-isomerase (EC 5.3.4.1) - alfalfa (clone GI) (fragments)
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 2742: from 469 to 488

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2743
- Ceres seq_id 1574061
- Location of start within SEQ ID NO 2741: at 22 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 2743: from 102 to 208 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1291
- gi No. 99991
- Description: protein disulfide-isomerase (EC 5.3.4.1) - alfalfa (clone GI) (fragments)
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 2743: from 462 to 481

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2744
- Ceres seq_id 1574062
- Location of start within SEQ ID NO 2741: at 238 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 2744: from 30 to 136 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1292
- gi No. 99991
- Description: protein disulfide-isomerase (EC 5.3.4.1) - alfalfa (clone GI) (fragments)
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 2744: from 390 to 409

Maximum Length Sequence:
related to:
Clone IDs:
21335
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2745
- Ceres seq_id 1574066

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2746
- Ceres seq_id 1574067
- Location of start within SEQ ID NO 2745: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 2746: from 72 to 231 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1293
- gi No. 4903018
- Description: (AB027507) ACE [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2746: from 1 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2747
- Ceres seq_id 1574068
- Location of start within SEQ ID NO 2745: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 2747: from 61 to 220 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1294
- gi No. 4903018
- Description: (AB027507) ACE [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2747: from 1 to 235

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2748
- Ceres seq_id 1574069
- Location of start within SEQ ID NO 2745: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 2748: from 37 to 196 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1295
- gi No. 4903018
- Description: (AB027507) ACE [Arabidopsis thaliana]
- % Identity: 99.6
- Alignment Length: 246
- Location of Alignment in SEQ ID NO 2748: from 1 to 211

Maximum Length Sequence:
related to:
Clone IDs:
6922
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2749
- Ceres seq_id 1574074

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2750
- Ceres seq_id 1574075
- Location of start within SEQ ID NO 2749: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 2750: from 123 to 165 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1296
- gi No. 2271465
- Description: (AF009563) 3-ketoacyl-CoA synthase [Brassica napus]
- % Identity: 76
- Alignment Length: 225
- Location of Alignment in SEQ ID NO 2750: from 1 to 224

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2751
- Ceres seq_id 1574076
- Location of start within SEQ ID NO 2749: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 2751: from 58 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1297
- gi No. 2271465
- Description: (AF009563) 3-ketoacyl-CoA synthase [Brassica napus]
- % Identity: 76
- Alignment Length: 225
- Location of Alignment in SEQ ID NO 2751: from 1 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2752
- Ceres seq_id 1574077
- Location of start within SEQ ID NO 2749: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 2752: from 29 to 71 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1298
- gi No. 2271465
- Description: (AF009563) 3-ketoacyl-CoA synthase [Brassica napus]
- % Identity: 76
- Alignment Length: 225
- Location of Alignment in SEQ ID NO 2752: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
33980
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 2753
- Ceres seq_id 1574086

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2754
- Ceres seq_id 1574087
- Location of start within SEQ ID NO 2753: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2754: from 68 to 136 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1299
- gi No. 5668775
- Description: (AC005916) Strong similarity to gb|Y18349 U2 snRNP auxiliary factor, small subunit from Oryza sativa. ESTs gb|AA586295 and gb|AA597332 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2754: from 11 to 291

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2755
- Ceres seq_id 1574088
- Location of start within SEQ ID NO 2753: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2755: from 23 to 91 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1300
- gi No. 5668775
- Description: (AC005916) Strong similarity to gb|Y18349 U2 snRNP auxiliary factor, small subunit from Oryza sativa. ESTs gb|AA586295 and gb|AA597332 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2755: from 1 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2756
- Ceres seq_id 1574089
- Location of start within SEQ ID NO 2753: at 155 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 2756: from 17 to 85 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1301
- gi No. 5668775
- Description: (AC005916) Strong similarity to gb|Y18349 U2 snRNP auxiliary factor, small subunit from Oryza sativa. ESTs gb|AA586295 and gb|AA597332 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2756: from 1 to 240

Maximum Length Sequence:
related to:
Clone IDs:
40916
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2757
- Ceres seq_id 1574093
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2758
- Ceres seq_id 1574094
- Location of start within SEQ ID NO 2757: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2758: from 152 to 192 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2759
- Ceres seq_id 1574095
- Location of start within SEQ ID NO 2757: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 2759: from 149 to 189 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
106244
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2760
- Ceres seq_id 1574100
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2761
- Ceres seq_id 1574101
- Location of start within SEQ ID NO 2760: at 157 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1302
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 57
- Location of Alignment in SEQ ID NO 2761: from 1 to 57

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2762
- Ceres seq_id 1574102
- Location of start within SEQ ID NO 2760: at 608 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1303
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 277
- Location of Alignment in SEQ ID NO 2762: from 1 to 174

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2763
- Ceres seq_id 1574103
- Location of start within SEQ ID NO 2760: at 665 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1304
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 97.5
- Alignment Length: 277
- Location of Alignment in SEQ ID NO 2763: from 1 to 155

Maximum Length Sequence:
related to:
Clone IDs:
156222
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2764
- Ceres seq_id 1574104
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2765
- Ceres seq_id 1574105
- Location of start within SEQ ID NO 2764: at 216 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1305
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2765: from 1 to 281

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2766
- Ceres seq_id 1574106
- Location of start within SEQ ID NO 2764: at 228 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1306
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2766: from 1 to 277

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2767
- Ceres seq_id 1574107
- Location of start within SEQ ID NO 2764: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1307
- gi No. 3935172
- Description: (AC004557) F17L21.15 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 281
- Location of Alignment in SEQ ID NO 2767: from 1 to 264

Maximum Length Sequence:
related to:
Clone IDs:
110252

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2768
- Ceres seq_id 1574122

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2769
- Ceres seq_id 1574123
- Location of start within SEQ ID NO 2768: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2769: from 180 to 411 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2770
- Ceres seq_id 1574124
- Location of start within SEQ ID NO 2768: at 567 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 2770: from 16 to 247 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
33790

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2771
- Ceres seq_id 1574133

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2772
- Ceres seq_id 1574134
- Location of start within SEQ ID NO 2771: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- O-methyltransferase
- Location within SEQ ID NO 2772: from 64 to 305 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1308
- gi No. 2781394
- Description: (U70424) O-methyltransferase 1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 2772: from 1 to 330

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2773
- Ceres seq_id 1574135
- Location of start within SEQ ID NO 2771: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- O-methyltransferase
- Location within SEQ ID NO 2773: from 50 to 291 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1309
- gi No. 2781394
- Description: (U70424) O-methyltransferase 1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 2773: from 1 to 316

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2774
- Ceres seq_id 1574136
- Location of start within SEQ ID NO 2771: at 242 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- O-methyltransferase
- Location within SEQ ID NO 2774: from 43 to 284 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1310
- gi No. 2781394
- Description: (U70424) O-methyltransferase 1 [Arabidopsis thaliana]
- % Identity: 99.7
- Alignment Length: 331
- Location of Alignment in SEQ ID NO 2774: from 1 to 309

Maximum Length Sequence:
related to:
Clone IDs:
23916
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2775
- Ceres seq_id 1574141
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2776
- Ceres seq_id 1574142
- Location of start within SEQ ID NO 2775: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1311
- gi No. 4914316
- Description: (AC005489) F14N23.2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 167
- Location of Alignment in SEQ ID NO 2776: from 24 to 190

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2777
- Ceres seq_id 1574143
- Location of start within SEQ ID NO 2775: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1312
- gi No. 4914316
- Description: (AC005489) F14N23.2 [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 167
- Location of Alignment in SEQ ID NO 2777: from 1 to 167

Maximum Length Sequence:
related to:
Clone IDs:
10026

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2778
- Ceres seq_id 1574148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2779
- Ceres seq_id 1574149
- Location of start within SEQ ID NO 2778: at 65 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L6e
- Location within SEQ ID NO 2779: from 58 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1313
- gi No. 464621
- Description: 60S RIBOSOMAL PROTEIN L6 (YL16-LIKE) >gi|280374|pir||S28586 ribosomal protein ML16 - common ice plant >gi|19539|emb|CAA49175| (X69378) ribosomal protein YL16 [Mesembryanthemum crystallinum]
- % Identity: 82.7
- Alignment Length: 231
- Location of Alignment in SEQ ID NO 2779: from 3 to 233

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2780
- Ceres seq_id 1574150
- Location of start within SEQ ID NO 2778: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L6e

- Location within SEQ ID NO 2780: from 31 to 206 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1314
- gi No. 464621
- Description: 60S RIBOSOMAL PROTEIN L6 (YL16-LIKE)
>gi|280374|pir||S28586 ribosomal protein ML16 - common ice plant
>gi|19539|emb|CAA49175| (X69378) ribosomal protein YL16 [Mesembryanthemum crystallinum]
- % Identity: 82.7
- Alignment Length: 231
- Location of Alignment in SEQ ID NO 2780: from 1 to 206

Maximum Length Sequence:
related to:
Clone IDs:
25911
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2781
- Ceres seq_id 1574151
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2782
- Ceres seq_id 1574152
- Location of start within SEQ ID NO 2781: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 1
- Location within SEQ ID NO 2782: from 1 to 264 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1315
- gi No. 1076442
- Description: beta-glucosidase (EC 3.2.1.21) - rape
>gi|757740|emb|CAA57913| (X82577) beta-glucosidase [Brassica napus]
- % Identity: 70.7
- Alignment Length: 259
- Location of Alignment in SEQ ID NO 2782: from 1 to 257

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2783
- Ceres seq_id 1574153
- Location of start within SEQ ID NO 2781: at 208 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 1
- Location within SEQ ID NO 2783: from 1 to 195 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1316
- gi No. 1076442
- Description: beta-glucosidase (EC 3.2.1.21) - rape
>gi|757740|emb|CAA57913| (X82577) beta-glucosidase [Brassica napus]
- % Identity: 70.7
- Alignment Length: 259
- Location of Alignment in SEQ ID NO 2783: from 1 to 188

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2784
- Ceres seq_id 1574154

- Location of start within SEQ ID NO 2781: at 292 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 1
- Location within SEQ ID NO 2784: from 1 to 167 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1317
- gi No. 1076442
- Description: beta-glucosidase (EC 3.2.1.21) - rape >gi|757740|emb|CAA57913| (X82577) beta-glucosidase [Brassica napus]
- % Identity: 70.7
- Alignment Length: 259
- Location of Alignment in SEQ ID NO 2784: from 1 to 160

Maximum Length Sequence:
related to:
Clone IDs:
30652
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2785
- Ceres seq_id 1574159
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2786
- Ceres seq_id 1574160
- Location of start within SEQ ID NO 2785: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type carbonic anhydrases
- Location within SEQ ID NO 2786: from 59 to 240 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1318
- gi No. 3176676
- Description: (AC003671) Similar to carbonic anhydrase gb|L19255 from Nicotiana tabacum. ESTs gb|AA597643, gb|T45390, gb|T43963 and gb|AA597734 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 258
- Location of Alignment in SEQ ID NO 2786: from 1 to 258

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2787
- Ceres seq_id 1574161
- Location of start within SEQ ID NO 2785: at 439 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prokaryotic-type carbonic anhydrases
- Location within SEQ ID NO 2787: from 1 to 127 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1319
- gi No. 3176676
- Description: (AC003671) Similar to carbonic anhydrase gb|L19255 from Nicotiana tabacum. ESTs gb|AA597643, gb|T45390, gb|T43963 and gb|AA597734 come from this gene. [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 258

- Location of Alignment in SEQ ID NO 2787: from 1 to 145

Maximum Length Sequence:
related to:
Clone IDs:
33965
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2788
- Ceres seq_id 1574162
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2789
- Ceres seq_id 1574163
- Location of start within SEQ ID NO 2788: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Prion protein
- Location within SEQ ID NO 2789: from 50 to 135 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1320
- gi No. 2832702
- Description: (AL021713) glycine-rich cell wall protein-like [Arabidopsis thaliana]
- % Identity: 98.5
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 2789: from 22 to 135

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2790
- Ceres seq_id 1574164
- Location of start within SEQ ID NO 2788: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2791
- Ceres seq_id 1574165
- Location of start within SEQ ID NO 2788: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2792
- Ceres seq_id 1574194
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2793
- Ceres seq_id 1574195
- Location of start within SEQ ID NO 2792: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Reverse transcriptase (RNA-dependent DNA polymerase)
- Location within SEQ ID NO 2793: from 5 to 75 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2794
- Ceres seq_id 1574197
- Location of start within SEQ ID NO 2792: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2795
- Ceres seq_id 1574203

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2796
- Ceres seq_id 1574204
- Location of start within SEQ ID NO 2795: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1321
- gi No. 4325366
- Description: (AF128396) similar to maize transposon MuDR-like proteins [Arabidopsis thaliana]
- % Identity: 85.7
- Alignment Length: 35
- Location of Alignment in SEQ ID NO 2796: from 1 to 34

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2797
- Ceres seq_id 1574205
- Location of start within SEQ ID NO 2795: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2798
- Ceres seq_id 1574218

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2799
- Ceres seq_id 1574219
- Location of start within SEQ ID NO 2798: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2800
- Ceres seq_id 1574220
- Location of start within SEQ ID NO 2798: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1322
- gi No. 2982435
- Description: (AL022224) retrotransposon like protein (fragment) [Arabidopsis thaliana]
- % Identity: 100
- Alignment Length: 35
- Location of Alignment in SEQ ID NO 2800: from 1 to 35

Maximum Length Sequence:
related to:
Clone IDs:
207829
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2801
- Ceres seq_id 1574260
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2802
- Ceres seq_id 1574261
- Location of start within SEQ ID NO 2801: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2803
- Ceres seq_id 1574262
- Location of start within SEQ ID NO 2801: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1323
- gi No. 1174867
- Description: UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C (UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 8.2 KD PROTEIN) >gi|633687|emb|CAA55862| (X79275) ubiquinol--cytochrome c reductase [Solanum tuberosum] [Solanum tuberosum]
- % Identity: 73.6
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 2803: from 1 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2804
- Ceres seq_id 1574263
- Location of start within SEQ ID NO 2801: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1324
- gi No. 1174867
- Description: UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C (UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 8.2 KD PROTEIN) >gi|633687|emb|CAA55862| (X79275) ubiquinol--cytochrome c reductase [Solanum tuberosum] [Solanum tuberosum]
- % Identity: 73.6
- Alignment Length: 72

- Location of Alignment in SEQ ID NO 2804: from 1 to 65

Maximum Length Sequence:
related to:
Clone IDs:
207834
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2805
- Ceres seq_id 1574264
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2806
- Ceres seq_id 1574265
- Location of start within SEQ ID NO 2805: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 2806: from 34 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1325
- gi No. 416731
- Description: POLLEN SPECIFIC PROTEIN C13 PRECURSOR >gi|82655|pir||JQ1107 18.3K protein precursor, pollen - maize >gi|255569|bbs|113677 (S44171) pollen specific protein [Zea mays=corn, Peptide, 170 aa] [Zea mays] >gi|1588669|prf||2209273A Zm13 [Zea mays]
- % Identity: 83.2
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 2806: from 1 to 110

Maximum Length Sequence:
related to:
Clone IDs:
208051
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2807
- Ceres seq_id 1574297
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2808
- Ceres seq_id 1574298
- Location of start within SEQ ID NO 2807: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 2808: from 3 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1326
- gi No. 2267597
- Description: (AF009413) 10 kDa chaperonin [Oryza sativa]
- % Identity: 88.9
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 2808: from 1 to 36

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2809
- Ceres seq_id 1574299
- Location of start within SEQ ID NO 2807: at 213 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
208141
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2810
- Ceres seq_id 1574312
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2811
- Ceres seq_id 1574313
- Location of start within SEQ ID NO 2810: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 2811: from 48 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1327
- gi No. 4263519
- Description: (AC004044) small nuclear riboprotein Sm-D1 [Arabidopsis thaliana]
- % Identity: 91.4
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 2811: from 44 to 158

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2812
- Ceres seq_id 1574314
- Location of start within SEQ ID NO 2810: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 2812: from 5 to 70 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1328
- gi No. 4263519
- Description: (AC004044) small nuclear riboprotein Sm-D1 [Arabidopsis thaliana]
- % Identity: 91.4
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 2812: from 1 to 115

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2813
- Ceres seq_id 1574315
- Location of start within SEQ ID NO 2810: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 2813: from 1 to 63 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1329

- gi No. 4263519
- Description: (AC004044) small nuclear riboprotein Sm-D1 [Arabidopsis thaliana]
- % Identity: 91.4
- Alignment Length: 116
- Location of Alignment in SEQ ID NO 2813: from 1 to 108

Maximum Length Sequence:
related to:
Clone IDs:
208722
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2814
- Ceres seq_id 1574376
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2815
- Ceres seq_id 1574377
- Location of start within SEQ ID NO 2814: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bowman-Birk serine protease inhibitor family
- Location within SEQ ID NO 2815: from 64 to 120 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1330
- gi No. 400077
- Description: BOWMAN-BIRK TYPE WOUND INDUCED PROTEINASE INHIBITOR WIP1 PRECURSOR >gi|486850|pir||S36236 wound-induced protein WIP1 precursor - maize >gi|296180|emb|CAA50519| (X71396) wound induced protein [Zea mays]
- % Identity: 99
- Alignment Length: 102
- Location of Alignment in SEQ ID NO 2815: from 19 to 120

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2816
- Ceres seq_id 1574378
- Location of start within SEQ ID NO 2814: at 57 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bowman-Birk serine protease inhibitor family
- Location within SEQ ID NO 2816: from 46 to 102 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1331
- gi No. 400077
- Description: BOWMAN-BIRK TYPE WOUND INDUCED PROTEINASE INHIBITOR WIP1 PRECURSOR >gi|486850|pir||S36236 wound-induced protein WIP1 precursor - maize >gi|296180|emb|CAA50519| (X71396) wound induced protein [Zea mays]
- % Identity: 99
- Alignment Length: 102
- Location of Alignment in SEQ ID NO 2816: from 1 to 102

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2817
- Ceres seq_id 1574379
- Location of start within SEQ ID NO 2814: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Bowman-Birk serine protease inhibitor family
- Location within SEQ ID NO 2817: from 28 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1332
- gi No. 400077
- Description: BOWMAN-BIRK TYPE WOUND INDUCED PROTEINASE INHIBITOR WIP1 PRECURSOR >gi|486850|pir||S36236 wound-induced protein WIP1 precursor - maize >gi|296180|emb|CAA50519| (X71396) wound induced protein [Zea mays]
- % Identity: 99
- Alignment Length: 102
- Location of Alignment in SEQ ID NO 2817: from 1 to 84

Maximum Length Sequence:
related to:
Clone IDs:
208741
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2818
- Ceres seq_id 1574380

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2819
- Ceres seq_id 1574381
- Location of start within SEQ ID NO 2818: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 2819: from 37 to 137 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2820
- Ceres seq_id 1574382
- Location of start within SEQ ID NO 2818: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 2820: from 7 to 107 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
208755
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2821
- Ceres seq_id 1574386

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2822
- Ceres seq_id 1574387
- Location of start within SEQ ID NO 2821: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1333
- gi No. 2811026

- Description: TCP1-CHAPERONIN COFACTOR A HOMOLOG >gi|1946375 (U93215) TCP1-chaperonin cofactor A isolog [Arabidopsis thaliana] >gi|2347204 (AC002338) TCP1-chaperonin cofactor A isolog [Arabidopsis thaliana]
- % Identity: 87.5
- Alignment Length: 80
- Location of Alignment in SEQ ID NO 2822: from 36 to 115

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2823
- Ceres seq_id 1574388
- Location of start within SEQ ID NO 2821: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2824
- Ceres seq_id 1574389
- Location of start within SEQ ID NO 2821: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1334
- gi No. 2811026
- Description: TCP1-CHAPERONIN COFACTOR A HOMOLOG >gi|1946375 (U93215) TCP1-chaperonin cofactor A isolog [Arabidopsis thaliana] >gi|2347204 (AC002338) TCP1-chaperonin cofactor A isolog [Arabidopsis thaliana]
- % Identity: 87.5
- Alignment Length: 80
- Location of Alignment in SEQ ID NO 2824: from 1 to 80

Maximum Length Sequence:
related to:
Clone IDs:
208863
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2825
- Ceres seq_id 1574403

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2826
- Ceres seq_id 1574404
- Location of start within SEQ ID NO 2825: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2827
- Ceres seq_id 1574405
- Location of start within SEQ ID NO 2825: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1335
- gi No. 4039153

- Description: (AF104221) low temperature and salt responsive protein LTI6A [Arabidopsis thaliana] >gi|4325217|gb|AAD17302| (AF122005) hydrophobic protein [Arabidopsis thaliana]
- % Identity: 75.5
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 2827: from 47 to 95

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2828
- Ceres seq_id 1574406
- Location of start within SEQ ID NO 2825: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1336
- gi No. 4039153
- Description: (AF104221) low temperature and salt responsive protein LTI6A [Arabidopsis thaliana] >gi|4325217|gb|AAD17302| (AF122005) hydrophobic protein [Arabidopsis thaliana]
- % Identity: 75.5
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 2828: from 8 to 56

Maximum Length Sequence:
related to:
Clone IDs:
208992

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2829
- Ceres seq_id 1574421

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2830
- Ceres seq_id 1574422
- Location of start within SEQ ID NO 2829: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2830: from 7 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1337
- gi No. 4127352
- Description: (AJ010451) glutathione transferase [Alopecurus myosuroides]
- % Identity: 74.1
- Alignment Length: 112
- Location of Alignment in SEQ ID NO 2830: from 4 to 115

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2831
- Ceres seq_id 1574423
- Location of start within SEQ ID NO 2829: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2831: from 1 to 94 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1338
- gi No. 4127352
- Description: (AJ010451) glutathione transferase [Alopecurus myosuroides]
- % Identity: 74.1
- Alignment Length: 112
- Location of Alignment in SEQ ID NO 2831: from 1 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2832
- Ceres seq_id 1574424
- Location of start within SEQ ID NO 2829: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2832: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1339
- gi No. 4127352
- Description: (AJ010451) glutathione transferase [Alopecurus myosuroides]
- % Identity: 74.1
- Alignment Length: 112
- Location of Alignment in SEQ ID NO 2832: from 1 to 74

Maximum Length Sequence:
related to:
Clone IDs:
209169
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2833
- Ceres seq_id 1574438

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2834
- Ceres seq_id 1574439
- Location of start within SEQ ID NO 2833: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2835
- Ceres seq_id 1574440
- Location of start within SEQ ID NO 2833: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S13
- Location within SEQ ID NO 2835: from 14 to 67 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1340
- gi No. 464707
- Description: 40S RIBOSOMAL PROTEIN S18 >gi|480908|pir||S37496 ribosomal protein S18.A - Arabidopsis thaliana >gi|405613|emb|CAA80684| (Z23165) ribosomal protein S18A [Arabidopsis thaliana] >gi|434343|emb|CAA82273| (Z28701) S18
- % Identity: 77.6

- Alignment Length: 67
- Location of Alignment in SEQ ID NO 2835: from 1 to 67

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2836
- Ceres seq_id 1574441
- Location of start within SEQ ID NO 2833: at 171 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1341
- gi No. 464707
- Description: 40S RIBOSOMAL PROTEIN S18 >gi|480908|pir||S37496 ribosomal protein S18.A - Arabidopsis thaliana >gi|405613|emb|CAA80684| (Z23165) ribosomal protein S18A [Arabidopsis thaliana] >gi|434343|emb|CAA82273| (Z28701) S18
- % Identity: 77.6
- Alignment Length: 67
- Location of Alignment in SEQ ID NO 2836: from 1 to 41

Maximum Length Sequence:
related to:
Clone IDs:
209239
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2837
- Ceres seq_id 1574446
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2838
- Ceres seq_id 1574447
- Location of start within SEQ ID NO 2837: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1342
- gi No. 132736
- Description: 60S RIBOSOMAL PROTEIN L19 (VEGETATIVE SPECIFIC PROTEIN V14) (22 KD CALMODULIN-BINDING PROTEIN) >gi|71266|pir||R5DO9E ribosomal protein L19.e - slime mold (Dictyostelium discoideum) >gi|295737|emb|CAA33443| (X15383) V14
- % Identity: 73.7
- Alignment Length: 38
- Location of Alignment in SEQ ID NO 2838: from 37 to 74

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2839
- Ceres seq_id 1574448
- Location of start within SEQ ID NO 2837: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1343
- gi No. 132736
- Description: 60S RIBOSOMAL PROTEIN L19 (VEGETATIVE SPECIFIC PROTEIN V14) (22 KD CALMODULIN-BINDING PROTEIN) >gi|71266|pir||R5DO9E ribosomal protein L19.e - slime mold (Dictyostelium discoideum) >gi|295737|emb|CAA33443| (X15383) V14
- % Identity: 73.7

- Alignment Length: 38
- Location of Alignment in SEQ ID NO 2839: from 1 to 38

Maximum Length Sequence:
related to:
Clone IDs:
210209
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2840
- Ceres seq_id 1574509
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2841
- Ceres seq_id 1574510
- Location of start within SEQ ID NO 2840: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 2841: from 58 to 192 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1344
- gi No. 2829742
- Description: 40S RIBOSOMAL PROTEIN S23 >gi|1754684 (U81008) ribosomal protein S23 [Brugia malayi]
- % Identity: 78.3
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 2841: from 52 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2842
- Ceres seq_id 1574511
- Location of start within SEQ ID NO 2840: at 156 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 2842: from 7 to 141 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1345
- gi No. 2829742
- Description: 40S RIBOSOMAL PROTEIN S23 >gi|1754684 (U81008) ribosomal protein S23 [Brugia malayi]
- % Identity: 78.3
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 2842: from 1 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2843
- Ceres seq_id 1574512
- Location of start within SEQ ID NO 2840: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 2843: from 1 to 135 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1346
- gi No. 2829742

- Description: 40S RIBOSOMAL PROTEIN S23 >gi|1754684 (U81008) ribosomal protein S23 [Brugia malayi]
- % Identity: 78.3
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 2843: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
210468
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2844
- Ceres seq_id 1574523

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2845
- Ceres seq_id 1574524
- Location of start within SEQ ID NO 2844: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2845: from 33 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1347
- gi No. 4468792
- Description: (AJ010295) Glutathione transferase III(a) [Zea mays]
- % Identity: 81.9
- Alignment Length: 94
- Location of Alignment in SEQ ID NO 2845: from 30 to 122

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2846
- Ceres seq_id 1574525
- Location of start within SEQ ID NO 2844: at 89 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2846: from 4 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1348
- gi No. 4468792
- Description: (AJ010295) Glutathione transferase III(a) [Zea mays]
- % Identity: 81.9
- Alignment Length: 94
- Location of Alignment in SEQ ID NO 2846: from 1 to 93

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2847
- Ceres seq_id 1574526
- Location of start within SEQ ID NO 2844: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2847: from 1 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1349

- gi No. 4468792
- Description: (AJ010295) Glutathione transferase III(a) [Zea mays]
- % Identity: 81.9
- Alignment Length: 94
- Location of Alignment in SEQ ID NO 2847: from 1 to 85

Maximum Length Sequence:
related to:
Clone IDs:
210918
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2848
- Ceres seq_id 1574547

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2849
- Ceres seq_id 1574548
- Location of start within SEQ ID NO 2848: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1350
- gi No. 1350890
- Description: 30S RIBOSOMAL PROTEIN S31, CHLOROPLAST >gi|477193|pir||A48391 ribosomal protein S31 - spinach (fragment) >gi|300415|bbs|132204 ribosomal protein S31 [Spinacia oleracea=spinach, cv Alwaro, leaves, Peptide Chloroplast Partial, 43 aa]
- % Identity: 77.4
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 2849: from 94 to 124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2850
- Ceres seq_id 1574549
- Location of start within SEQ ID NO 2848: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 2850: from 4 to 84 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2851
- Ceres seq_id 1574550
- Location of start within SEQ ID NO 2848: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1351
- gi No. 1350890
- Description: 30S RIBOSOMAL PROTEIN S31, CHLOROPLAST >gi|477193|pir||A48391 ribosomal protein S31 - spinach (fragment) >gi|300415|bbs|132204 ribosomal protein S31 [Spinacia oleracea=spinach, cv Alwaro, leaves, Peptide Chloroplast Partial, 43 aa]
- % Identity: 77.4
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 2851: from 69 to 99

Maximum Length Sequence:
related to:
Clone IDs:
211358
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2852
- Ceres seq_id 1574572
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2853
- Ceres seq_id 1574573
- Location of start within SEQ ID NO 2852: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 2853: from 4 to 90 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2854
- Ceres seq_id 1574574
- Location of start within SEQ ID NO 2852: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2855
- Ceres seq_id 1574575
- Location of start within SEQ ID NO 2852: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 2855: from 1 to 57 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
211416
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2856
- Ceres seq_id 1574580
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2857
- Ceres seq_id 1574581
- Location of start within SEQ ID NO 2856: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 2857: from 3 to 49 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2858

- Ceres seq_id 1574582
- Location of start within SEQ ID NO 2856: at 258 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
211830
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2859
- Ceres seq_id 1574616
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2860
- Ceres seq_id 1574617
- Location of start within SEQ ID NO 2859: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 2860: from 1 to 78 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2861
- Ceres seq_id 1574618
- Location of start within SEQ ID NO 2859: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Major intrinsic protein
- Location within SEQ ID NO 2861: from 81 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1352
- gi No. 4584429
- Description: (AJ237751) aquaglyceroporin [Nicotiana tabacum]
- % Identity: 70.3
- Alignment Length: 74
- Location of Alignment in SEQ ID NO 2861: from 80 to 153

Maximum Length Sequence:
related to:
Clone IDs:
211845
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2862
- Ceres seq_id 1574619
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2863
- Ceres seq_id 1574620
- Location of start within SEQ ID NO 2862: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2863: from 31 to 109 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1353
- gi No. 2497909
- Description: METALLOTHIONEIN-LIKE PROTEIN 1B (MT-1B) >gi|1361996|pir||S57859 metallothionein 1b - Arabidopsis thaliana >gi|1086459 (U11254) metallothionein [Arabidopsis thaliana]
- % Identity: 73.3
- Alignment Length: 15
- Location of Alignment in SEQ ID NO 2863: from 38 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2864
- Ceres seq_id 1574621
- Location of start within SEQ ID NO 2862: at 93 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2864: from 1 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1354
- gi No. 2497909
- Description: METALLOTHIONEIN-LIKE PROTEIN 1B (MT-1B) >gi|1361996|pir||S57859 metallothionein 1b - Arabidopsis thaliana >gi|1086459 (U11254) metallothionein [Arabidopsis thaliana]
- % Identity: 73.3
- Alignment Length: 15
- Location of Alignment in SEQ ID NO 2864: from 8 to 22

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2865
- Ceres seq_id 1574622
- Location of start within SEQ ID NO 2862: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2865: from 1 to 57 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
212067
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2866
- Ceres seq_id 1574626
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2867
- Ceres seq_id 1574627
- Location of start within SEQ ID NO 2866: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1355
- gi No. 1350526
- Description: (L47748) glycine-rich cell wall protein [Picea glauca]

- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2867: from 58 to 69

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2868
- Ceres seq_id 1574628
- Location of start within SEQ ID NO 2866: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2869
- Ceres seq_id 1574629
- Location of start within SEQ ID NO 2866: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
212187
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2870
- Ceres seq_id 1574636
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2871
- Ceres seq_id 1574637
- Location of start within SEQ ID NO 2870: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2872
- Ceres seq_id 1574638
- Location of start within SEQ ID NO 2870: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2873
- Ceres seq_id 1574639
- Location of start within SEQ ID NO 2870: at 210 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 2873: from 20 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1356
- gi No. 100883

- Description: heat shock protein 17.2 - maize
>gi|22335|emb|CAA46641| (X65725) heat shock protein 17.2 [Zea mays]
- % Identity: 83.6
- Alignment Length: 146
- Location of Alignment in SEQ ID NO 2873: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
213078
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2874
- Ceres seq_id 1574673
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2875
- Ceres seq_id 1574674
- Location of start within SEQ ID NO 2874: at 38 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2876
- Ceres seq_id 1574675
- Location of start within SEQ ID NO 2874: at 206 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2877
- Ceres seq_id 1574676
- Location of start within SEQ ID NO 2874: at 336 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1357
- gi No. 2621798
- Description: (AE000850) transcriptional regulator [Methanobacterium thermoautotrophicum]
- % Identity: 75
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 2877: from 88 to 111

Maximum Length Sequence:
related to:
Clone IDs:
213080
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2878
- Ceres seq_id 1574677
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2879
- Ceres seq_id 1574678
- Location of start within SEQ ID NO 2878: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1358
- gi No. 1617206
- Description: (Z72489) CP12 [Pisum sativum]
- % Identity: 81.3
- Alignment Length: 64
- Location of Alignment in SEQ ID NO 2879: from 76 to 139

Maximum Length Sequence:
related to:
Clone IDs:
214475
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2880
- Ceres seq_id 1574693
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2881
- Ceres seq_id 1574694
- Location of start within SEQ ID NO 2880: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2881: from 25 to 167 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1359
- gi No. 2511594
- Description: (Y13694) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana] >gi|2827525|emb|CAA16533| (AL021633) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana]
- % Identity: 72.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2881: from 20 to 242

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2882
- Ceres seq_id 1574695
- Location of start within SEQ ID NO 2880: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2882: from 4 to 146 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1360
- gi No. 2511594
- Description: (Y13694) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana] >gi|2827525|emb|CAA16533| (AL021633) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana]
- % Identity: 72.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2882: from 1 to 221

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2883
- Ceres seq_id 1574696
- Location of start within SEQ ID NO 2880: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2883: from 1 to 121 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1361
- gi No. 2511594
- Description: (Y13694) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana] >gi|2827525|emb|CAA16533| (AL021633) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana]
- % Identity: 72.6
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 2883: from 1 to 196

Maximum Length Sequence:
related to:
Clone IDs:
215389
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2884
- Ceres seq_id 1574701
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2885
- Ceres seq_id 1574702
- Location of start within SEQ ID NO 2884: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2886
- Ceres seq_id 1574703
- Location of start within SEQ ID NO 2884: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2886: from 6 to 197 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2887
- Ceres seq_id 1574704
- Location of start within SEQ ID NO 2884: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 2887: from 1 to 140 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
215916

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2888
- Ceres seq_id 1574713

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2889
- Ceres seq_id 1574714
- Location of start within SEQ ID NO 2888: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1362
- gi No. 4056457
- Description: (AC005990) ESTs gb|234051 and gb|F13722 come from this gene. [Arabidopsis thaliana]
- % Identity: 77.8
- Alignment Length: 90
- Location of Alignment in SEQ ID NO 2889: from 86 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2890
- Ceres seq_id 1574715
- Location of start within SEQ ID NO 2888: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1363
- gi No. 1084362
- Description: extensin-like protein S3 - alfalfa >gi|535586 (L36120) proline rich protein [Medicago sativa]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 2890: from 102 to 112

Maximum Length Sequence:
related to:
Clone IDs:
216117

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2891
- Ceres seq_id 1574735

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2892
- Ceres seq_id 1574736
- Location of start within SEQ ID NO 2891: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2893
- Ceres seq_id 1574737
- Location of start within SEQ ID NO 2891: at 187 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 2893: from 9 to 89 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
217321
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2894
- Ceres seq_id 1574738
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2895
- Ceres seq_id 1574739
- Location of start within SEQ ID NO 2894: at 181 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallo-beta-lactamase superfamily
- Location within SEQ ID NO 2895: from 3 to 125 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1364
- gi No. 3913733
- Description: HYDROXYACYLGLUTATHIONE HYDROLASE CYTOPLASMIC ISOZYME (GLYOXALASE II) (GLX II) >gi|1924921|emb|CAA69644| (Y08357) hydroxyacylglutathione hydrolase [Arabidopsis thaliana]
- % Identity: 71.2
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 2895: from 1 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2896
- Ceres seq_id 1574740
- Location of start within SEQ ID NO 2894: at 376 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallo-beta-lactamase superfamily
- Location within SEQ ID NO 2896: from 1 to 60 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1365
- gi No. 3913733
- Description: HYDROXYACYLGLUTATHIONE HYDROLASE CYTOPLASMIC ISOZYME (GLYOXALASE II) (GLX II) >gi|1924921|emb|CAA69644| (Y08357) hydroxyacylglutathione hydrolase [Arabidopsis thaliana]
- % Identity: 71.2
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 2896: from 1 to 60

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2897
- Ceres seq_id 1574741
- Location of start within SEQ ID NO 2894: at 636 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

217695
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2898
- Ceres seq_id 1574754
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2899
- Ceres seq_id 1574755
- Location of start within SEQ ID NO 2898: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2900
- Ceres seq_id 1574756
- Location of start within SEQ ID NO 2898: at 321 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1366
- gi No. 2104959
- Description: (U96925) immunophilin [Vicia faba]
- % Identity: 84.1
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 2900: from 1 to 37

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2901
- Ceres seq_id 1574757
- Location of start within SEQ ID NO 2898: at 327 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1367
- gi No. 2104959
- Description: (U96925) immunophilin [Vicia faba]
- % Identity: 84.1
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 2901: from 1 to 35

Maximum Length Sequence:
related to:
Clone IDs:
217731
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2902
- Ceres seq_id 1574758
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2903
- Ceres seq_id 1574759
- Location of start within SEQ ID NO 2902: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2903: from 26 to 105 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1368
- gi No. 2497904
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1667588 (U77294) metallothionein-like protein [Oryza sativa] >gi|2326785|emb|CAA69845| (Y08529) metallothionein-like protein [Oryza sativa] >gi|4097338 (U57638) metallothionein-like protein sativa]
- % Identity: 72.3
- Alignment Length: 85
- Location of Alignment in SEQ ID NO 2903: from 26 to 106

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2904
- Ceres seq_id 1574760
- Location of start within SEQ ID NO 2902: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2904: from 3 to 82 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1369
- gi No. 2497904
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1667588 (U77294) metallothionein-like protein [Oryza sativa] >gi|2326785|emb|CAA69845| (Y08529) metallothionein-like protein [Oryza sativa] >gi|4097338 (U57638) metallothionein-like protein sativa]
- % Identity: 72.3
- Alignment Length: 85
- Location of Alignment in SEQ ID NO 2904: from 3 to 83

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2905
- Ceres seq_id 1574761
- Location of start within SEQ ID NO 2902: at 145 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 2905: from 1 to 57 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1370
- gi No. 2497904
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1667588 (U77294) metallothionein-like protein [Oryza sativa] >gi|2326785|emb|CAA69845| (Y08529) metallothionein-like protein [Oryza sativa] >gi|4097338 (U57638) metallothionein-like protein sativa]
- % Identity: 72.3
- Alignment Length: 85
- Location of Alignment in SEQ ID NO 2905: from 1 to 58

Maximum Length Sequence:
related to:
Clone IDs:
217737
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2906
- Ceres seq_id 1574762

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2907

- Ceres seq_id 1574763
- Location of start within SEQ ID NO 2906: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 2907: from 53 to 179 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1371
- gi No. 231509
- Description: ACTIN DEPOLYMERIZING FACTOR (ADF) >gi|419809|pir||S30935 actin-depolymerizing factor - trumpet lily >gi|22748|emb|CAA78483| (Z14110) actin depolymerizing factor [Lilium longiflorum]
- % Identity: 84.8
- Alignment Length: 138
- Location of Alignment in SEQ ID NO 2907: from 42 to 179

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2908
- Ceres seq_id 1574764
- Location of start within SEQ ID NO 2906: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2909
- Ceres seq_id 1574765
- Location of start within SEQ ID NO 2906: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cofilin/tropomyosin-type actin-binding proteins
- Location within SEQ ID NO 2909: from 12 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1372
- gi No. 231509
- Description: ACTIN DEPOLYMERIZING FACTOR (ADF) >gi|419809|pir||S30935 actin-depolymerizing factor - trumpet lily >gi|22748|emb|CAA78483| (Z14110) actin depolymerizing factor [Lilium longiflorum]
- % Identity: 84.8
- Alignment Length: 138
- Location of Alignment in SEQ ID NO 2909: from 1 to 138

Maximum Length Sequence:
related to:
Clone IDs:
217831

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2910
- Ceres seq_id 1574769

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2911
- Ceres seq_id 1574770
- Location of start within SEQ ID NO 2910: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2912
- Ceres seq_id 1574771
- Location of start within SEQ ID NO 2910: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2913
- Ceres seq_id 1574772
- Location of start within SEQ ID NO 2910: at 340 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Retroviral aspartyl proteases
- Location within SEQ ID NO 2913: from 1 to 52 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
217837
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2914
- Ceres seq_id 1574773

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2915
- Ceres seq_id 1574774
- Location of start within SEQ ID NO 2914: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1373
- gi No. 322759
- Description: pistil extensin-like protein (clone pMG14) - common tobacco (fragment) >gi|19927|emb|CAA78393| (Z14015) pistil extensin like protein [Nicotiana tabacum]
- % Identity: 76.5
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 2915: from 7 to 22

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2916
- Ceres seq_id 1574775
- Location of start within SEQ ID NO 2914: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Copper binding proteins, plastocyanin/azurin family
- Location within SEQ ID NO 2916: from 92 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1374

- gi No. 3885896
- Description: (AF093636) plastocyanin precursor [Oryza sativa]
- % Identity: 90
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 2916: from 92 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2917
- Ceres seq_id 1574776
- Location of start within SEQ ID NO 2914: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Copper binding proteins, plastocyanin/azurin family
- Location within SEQ ID NO 2917: from 86 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1375
- gi No. 3885896
- Description: (AF093636) plastocyanin precursor [Oryza sativa]
- % Identity: 90
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 2917: from 86 to 155

Maximum Length Sequence:
related to:
Clone IDs:
218125
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2918
- Ceres seq_id 1574797

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2919
- Ceres seq_id 1574798
- Location of start within SEQ ID NO 2918: at 43 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 2919: from 36 to 99 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218158
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2920
- Ceres seq_id 1574803

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2921
- Ceres seq_id 1574804
- Location of start within SEQ ID NO 2920: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 2921: from 16 to 68 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2922
- Ceres seq_id 1574805
- Location of start within SEQ ID NO 2920: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2923
- Ceres seq_id 1574806
- Location of start within SEQ ID NO 2920: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218216
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2924
- Ceres seq_id 1574814
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2925
- Ceres seq_id 1574815
- Location of start within SEQ ID NO 2924: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2925: from 80 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1376
- gi No. 3421087
- Description: (AF043524) 20S proteasome subunit PAE1 [Arabidopsis thaliana]
- % Identity: 95.3
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 2925: from 45 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2926
- Ceres seq_id 1574816
- Location of start within SEQ ID NO 2924: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2926: from 36 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1377
- gi No. 3421087
- Description: (AF043524) 20S proteasome subunit PAE1 [Arabidopsis thaliana]
- % Identity: 95.3

- Alignment Length: 150
- Location of Alignment in SEQ ID NO 2926: from 1 to 149

Maximum Length Sequence:
related to:
Clone IDs:
218224
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2927
- Ceres seq_id 1574817
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2928
- Ceres seq_id 1574818
- Location of start within SEQ ID NO 2927: at 214 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 2928: from 14 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218246
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2929
- Ceres seq_id 1574819
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2930
- Ceres seq_id 1574820
- Location of start within SEQ ID NO 2929: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2931
- Ceres seq_id 1574821
- Location of start within SEQ ID NO 2929: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pollen proteins Ole e I family
- Location within SEQ ID NO 2931: from 33 to 145 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
241409
218404
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2932
- Ceres seq_id 1574838
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2933
- Ceres seq_id 1574839

- Location of start within SEQ ID NO 2932: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sperm histone P2
- Location within SEQ ID NO 2933: from 3 to 97 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2934
- Ceres seq_id 1574840
- Location of start within SEQ ID NO 2932: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1378
- gi No. 132918
- Description: 50S RIBOSOMAL PROTEIN L35, CHLOROPLAST PRECURSOR (CL35) >gi|81486|pir||A36107 ribosomal protein L35 precursor, chloroplast - spinach >gi|170139 (M60449) ribosomal protein L35 [Spinacia oleracea]
- % Identity: 71.8
- Alignment Length: 71
- Location of Alignment in SEQ ID NO 2934: from 93 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2935
- Ceres seq_id 1574841
- Location of start within SEQ ID NO 2932: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218405
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2936
- Ceres seq_id 1574842
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2937
- Ceres seq_id 1574843
- Location of start within SEQ ID NO 2936: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2938
- Ceres seq_id 1574844
- Location of start within SEQ ID NO 2936: at 112 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 2938: from 2 to 79 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1379
- gi No. 2407800
- Description: (Y12575) histone H2A.F/Z [Arabidopsis thaliana]
- % Identity: 79.7
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 2938: from 1 to 79

Maximum Length Sequence:
related to:
Clone IDs:
218421
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2939
- Ceres seq_id 1574849
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2940
- Ceres seq_id 1574850
- Location of start within SEQ ID NO 2939: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1380
- gi No. 1110587
- Description: (S79410) nuclear localization signals (NLS)-binding protein=spot-1 [mice, Balb/c, testes, Peptide, 140 aa] [Mus sp.]
- % Identity: 94.1
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 2940: from 6 to 22

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2941
- Ceres seq_id 1574852
- Location of start within SEQ ID NO 2939: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1381
- gi No. 1110587
- Description: (S79410) nuclear localization signals (NLS)-binding protein=spot-1 [mice, Balb/c, testes, Peptide, 140 aa] [Mus sp.]
- % Identity: 81
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 2941: from 1 to 21

Maximum Length Sequence:
related to:
Clone IDs:
218458
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2942
- Ceres seq_id 1574853
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2943
- Ceres seq_id 1574854
- Location of start within SEQ ID NO 2942: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Glutathione S-transferases.
- Location within SEQ ID NO 2943: from 61 to 237 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2944
- Ceres seq_id 1574855
- Location of start within SEQ ID NO 2942: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 2944: from 14 to 190 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218566
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2945
- Ceres seq_id 1574868
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2946
- Ceres seq_id 1574869
- Location of start within SEQ ID NO 2945: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 2946: from 6 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1382
- gi No. 1498055
- Description: (U64437) novel protein [Zea mays]
- % Identity: 99.2
- Alignment Length: 118
- Location of Alignment in SEQ ID NO 2946: from 1 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2947
- Ceres seq_id 1574870
- Location of start within SEQ ID NO 2945: at 209 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- C2 domain
- Location within SEQ ID NO 2947: from 1 to 62 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1383
- gi No. 1498055
- Description: (U64437) novel protein [Zea mays]
- % Identity: 99.2
- Alignment Length: 118
- Location of Alignment in SEQ ID NO 2947: from 1 to 94

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 2948
- Ceres seq_id 1574871
- Location of start within SEQ ID NO 2945: at 289 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218701
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2949
- Ceres seq_id 1574880
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2950
- Ceres seq_id 1574881
- Location of start within SEQ ID NO 2949: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L9
- Location within SEQ ID NO 2950: from 32 to 99 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2951
- Ceres seq_id 1574882
- Location of start within SEQ ID NO 2949: at 353 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2952
- Ceres seq_id 1574883
- Location of start within SEQ ID NO 2949: at 362 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
218771
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2953
- Ceres seq_id 1574896
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2954
- Ceres seq_id 1574897
- Location of start within SEQ ID NO 2953: at 146 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2954: from 3 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1384
- gi No. 2511590
- Description: (Y13692) multicatalytic endopeptidase complex, proteasome component, beta subunit [Arabidopsis thaliana] >gi|3421111 (AF043534) 20S proteasome beta subunit PBD1 [Arabidopsis thaliana]
- % Identity: 76.2
- Alignment Length: 202
- Location of Alignment in SEQ ID NO 2954: from 1 to 202

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2955
- Ceres seq_id 1574898
- Location of start within SEQ ID NO 2953: at 251 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Proteasome A-type and B-type
- Location within SEQ ID NO 2955: from 1 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1385
- gi No. 2511590
- Description: (Y13692) multicatalytic endopeptidase complex, proteasome component, beta subunit [Arabidopsis thaliana] >gi|3421111 (AF043534) 20S proteasome beta subunit PBD1 [Arabidopsis thaliana]
- % Identity: 76.2
- Alignment Length: 202
- Location of Alignment in SEQ ID NO 2955: from 1 to 167

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2956
- Ceres seq_id 1574899
- Location of start within SEQ ID NO 2953: at 437 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1386
- gi No. 2511590
- Description: (Y13692) multicatalytic endopeptidase complex, proteasome component, beta subunit [Arabidopsis thaliana] >gi|3421111 (AF043534) 20S proteasome beta subunit PBD1 [Arabidopsis thaliana]
- % Identity: 76.2
- Alignment Length: 202
- Location of Alignment in SEQ ID NO 2956: from 1 to 105

Maximum Length Sequence:
related to:
Clone IDs:
218887

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2957
- Ceres seq_id 1574910

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2958
- Ceres seq_id 1574911
- Location of start within SEQ ID NO 2957: at 206 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 2958: from 48 to 150 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2959
- Ceres seq_id 1574912
- Location of start within SEQ ID NO 2957: at 401 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 2959: from 1 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
219406
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2960
- Ceres seq_id 1574947
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2961
- Ceres seq_id 1574948
- Location of start within SEQ ID NO 2960: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2962
- Ceres seq_id 1574949
- Location of start within SEQ ID NO 2960: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 2962: from 120 to 183 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2963
- Ceres seq_id 1574950
- Location of start within SEQ ID NO 2960: at 345 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 2963: from 6 to 69 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
219805
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2964
- Ceres seq_id 1574981
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2965
- Ceres seq_id 1574982
- Location of start within SEQ ID NO 2964: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dual specificity phosphatase, catalytic domain
- Location within SEQ ID NO 2965: from 55 to 183 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2966
- Ceres seq_id 1574983
- Location of start within SEQ ID NO 2964: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dual specificity phosphatase, catalytic domain
- Location within SEQ ID NO 2966: from 31 to 159 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
219894
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2967
- Ceres seq_id 1574984
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2968
- Ceres seq_id 1574985
- Location of start within SEQ ID NO 2967: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2969
- Ceres seq_id 1574986
- Location of start within SEQ ID NO 2967: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1387
- gi No. 3108209
- Description: (AF028809) eukaryotic cap-binding protein [Arabidopsis thaliana]
- % Identity: 78.8
- Alignment Length: 179
- Location of Alignment in SEQ ID NO 2969: from 35 to 213

Maximum Length Sequence:
related to:
Clone IDs:
220078
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2970
- Ceres seq_id 1574999
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2971
- Ceres seq_id 1575000
- Location of start within SEQ ID NO 2970: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1388
- gi No. 3786418
- Description: (AF098501) No definition line found [Caenorhabditis elegans]
- % Identity: 73.3
- Alignment Length: 16
- Location of Alignment in SEQ ID NO 2971: from 48 to 62

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2972
- Ceres seq_id 1575001
- Location of start within SEQ ID NO 2970: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2973
- Ceres seq_id 1575002
- Location of start within SEQ ID NO 2970: at 25 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1389
- gi No. 3786418
- Description: (AF098501) No definition line found [Caenorhabditis elegans]
- % Identity: 73.3
- Alignment Length: 16
- Location of Alignment in SEQ ID NO 2973: from 40 to 54

Maximum Length Sequence:
related to:
Clone IDs:
220098
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2974
- Ceres seq_id 1575003
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2975
- Ceres seq_id 1575004
- Location of start within SEQ ID NO 2974: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AMP-binding enzyme
- Location within SEQ ID NO 2975: from 1 to 112 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2976
- Ceres seq_id 1575005
- Location of start within SEQ ID NO 2974: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sperm histone P2
- Location within SEQ ID NO 2976: from 37 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1390
- gi No. 1170395
- Description: SPERM PROTAMINE P1 >gi|598425 (L35451) protamine P1 [Macropus agilis] >gi|1582117|prf||2117429K protamine P1 [Macropus agilis]
- % Identity: 76.9
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 2976: from 62 to 74

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2977
- Ceres seq_id 1575006
- Location of start within SEQ ID NO 2974: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AMP-binding enzyme
- Location within SEQ ID NO 2977: from 1 to 72 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
220148
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2978
- Ceres seq_id 1575013
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2979
- Ceres seq_id 1575014
- Location of start within SEQ ID NO 2978: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2980
- Ceres seq_id 1575015
- Location of start within SEQ ID NO 2978: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2981
- Ceres seq_id 1575016
- Location of start within SEQ ID NO 2978: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1391
- gi No. 2500378
- Description: 60S RIBOSOMAL PROTEIN L37
- % Identity: 77.7
- Alignment Length: 94
- Location of Alignment in SEQ ID NO 2981: from 1 to 94

Maximum Length Sequence:
related to:
Clone IDs:
220160
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2982
- Ceres seq_id 1575017
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2983
- Ceres seq_id 1575018
- Location of start within SEQ ID NO 2982: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 2983: from 108 to 175 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2984
- Ceres seq_id 1575019
- Location of start within SEQ ID NO 2982: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
220232
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2985
- Ceres seq_id 1575027
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2986
- Ceres seq_id 1575028
- Location of start within SEQ ID NO 2985: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2987
- Ceres seq_id 1575029
- Location of start within SEQ ID NO 2985: at 257 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2988
- Ceres seq_id 1575030
- Location of start within SEQ ID NO 2985: at 701 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AP endonuclease family 1
- Location within SEQ ID NO 2988: from 1 to 131 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
220347
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2989
- Ceres seq_id 1575038
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2990
- Ceres seq_id 1575039
- Location of start within SEQ ID NO 2989: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sperm histone P2
- Location within SEQ ID NO 2990: from 32 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1392
- gi No. 399783
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN (CLONE W10-1) >gi|388257|emb|CAA39222| (X55693) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 2990: from 30 to 41

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2991
- Ceres seq_id 1575040
- Location of start within SEQ ID NO 2989: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1393
- gi No. 3941737
- Description: (AF109719) BAT2 [Mus musculus]
- % Identity: 72.7
- Alignment Length: 11

- Location of Alignment in SEQ ID NO 2991: from 64 to 74

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2992
- Ceres seq_id 1575041
- Location of start within SEQ ID NO 2989: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1394
- gi No. 1362586
- Description: spermatid-specific protein T1 - longfin squid >gi|1042188|bbs|169195 T1=testis-specific pro-protamine [Loligo pealeii=squids, testis chromatin, Peptide, 79 aa]
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 2992: from 3 to 16

Maximum Length Sequence:
related to:
Clone IDs:
220593
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2993
- Ceres seq_id 1575056
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2994
- Ceres seq_id 1575057
- Location of start within SEQ ID NO 2993: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1395
- gi No. 4337178
- Description: (AC006416) T31J12.5 [Arabidopsis thaliana]
- % Identity: 84.3
- Alignment Length: 89
- Location of Alignment in SEQ ID NO 2994: from 68 to 156

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2995
- Ceres seq_id 1575058
- Location of start within SEQ ID NO 2993: at 371 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1396
- gi No. 4337178
- Description: (AC006416) T31J12.5 [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 2995: from 28 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2996
- Ceres seq_id 1575059
- Location of start within SEQ ID NO 2993: at 410 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1397
- gi No. 4337178
- Description: (AC006416) T31J12.5 [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 2996: from 15 to 84

Maximum Length Sequence:
related to:
Clone IDs:
220782

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 2997
- Ceres seq_id 1575067

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2998
- Ceres seq_id 1575068
- Location of start within SEQ ID NO 2997: at 154 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Skp1 family
- Location within SEQ ID NO 2998: from 9 to 62 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1398
- gi No. 3068809
- Description: (AF059295) Skp1 homolog [Arabidopsis thaliana]
- % Identity: 77.8
- Alignment Length: 54
- Location of Alignment in SEQ ID NO 2998: from 9 to 62

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 2999
- Ceres seq_id 1575069
- Location of start within SEQ ID NO 2997: at 181 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Skp1 family
- Location within SEQ ID NO 2999: from 1 to 53 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1399
- gi No. 3068809
- Description: (AF059295) Skp1 homolog [Arabidopsis thaliana]
- % Identity: 77.8
- Alignment Length: 54
- Location of Alignment in SEQ ID NO 2999: from 1 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3000
- Ceres seq_id 1575070
- Location of start within SEQ ID NO 2997: at 238 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1400
- gi No. 3068809
- Description: (AF059295) Skp1 homolog [Arabidopsis thaliana]
- % Identity: 77.8
- Alignment Length: 54
- Location of Alignment in SEQ ID NO 3000: from 1 to 34

Maximum Length Sequence:
related to:
Clone IDs:
221396
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3001
- Ceres seq_id 1575105
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3002
- Ceres seq_id 1575106
- Location of start within SEQ ID NO 3001: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 3002: from 37 to 91 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3003
- Ceres seq_id 1575107
- Location of start within SEQ ID NO 3001: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3004
- Ceres seq_id 1575108
- Location of start within SEQ ID NO 3001: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Homeobox domain
- Location within SEQ ID NO 3004: from 1 to 44 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
221429
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3005
- Ceres seq_id 1575124
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3006
- Ceres seq_id 1575125
- Location of start within SEQ ID NO 3005: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1401
- gi No. 82091
- Description: hydroxyproline-rich glycoprotein - tomato (fragment) >gi|170454 (M11759) cell wall hydroxyproline-rich glycoprotein [Lycopersicon esculentum] >gi|224726|prf||1111324A glycoprotein,Hyp rich [Lycopersicon esculentum]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3006: from 22 to 32

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3007
- Ceres seq_id 1575126
- Location of start within SEQ ID NO 3005: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3008
- Ceres seq_id 1575127
- Location of start within SEQ ID NO 3005: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
221533
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3009
- Ceres seq_id 1575143
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3010
- Ceres seq_id 1575144
- Location of start within SEQ ID NO 3009: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 3010: from 28 to 50 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1402
- gi No. 1362015
- Description: zinc finger protein 1 - Arabidopsis thaliana >gi|2129779|pir||S71240 zinc finger protein 1 - Arabidopsis thaliana >gi|790673 (L39644) zinc finger protein [Arabidopsis thaliana] >gi|1297186 (U53501) zinc finger protein 1 [Arabidopsis thaliana]
- % Identity: 74.3
- Alignment Length: 35
- Location of Alignment in SEQ ID NO 3010: from 25 to 59

Maximum Length Sequence:
related to:
Clone IDs:
221824

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3011
- Ceres seq_id 1575163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3012
- Ceres seq_id 1575164
- Location of start within SEQ ID NO 3011: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 3012: from 14 to 172 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3013
- Ceres seq_id 1575165
- Location of start within SEQ ID NO 3011: at 152 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 3013: from 7 to 165 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3014
- Ceres seq_id 1575166
- Location of start within SEQ ID NO 3011: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Signal peptidase I
- Location within SEQ ID NO 3014: from 1 to 159 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222156

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3015
- Ceres seq_id 1575181

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3016
- Ceres seq_id 1575182
- Location of start within SEQ ID NO 3015: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3017
- Ceres seq_id 1575183
- Location of start within SEQ ID NO 3015: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase
- Location within SEQ ID NO 3017: from 41 to 101 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3018
- Ceres seq_id 1575184
- Location of start within SEQ ID NO 3015: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222227
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3019
- Ceres seq_id 1575191
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3020
- Ceres seq_id 1575192
- Location of start within SEQ ID NO 3019: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1403
- gi No. 729671
- Description: HISTONE H2A >gi|473603 (U08225) histone H2A [Zea mays]
- % Identity: 80.6
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 3020: from 27 to 57

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3021
- Ceres seq_id 1575193
- Location of start within SEQ ID NO 3019: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1404
- gi No. 729671
- Description: HISTONE H2A >gi|473603 (U08225) histone H2A [Zea mays]
- % Identity: 80.6
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 3021: from 4 to 34

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3022
- Ceres seq_id 1575194
- Location of start within SEQ ID NO 3019: at 236 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
303745
222258
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3023
- Ceres seq_id 1575195
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3024
- Ceres seq_id 1575196
- Location of start within SEQ ID NO 3023: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-III pyridoxal-phosphate
- Location within SEQ ID NO 3024: from 1 to 92 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1405
- gi No. 5731259
- Description: (AF166351) alanine:glyoxylate aminotransferase 2 homolog [Arabidopsis thaliana]
- % Identity: 77
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 3024: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3025
- Ceres seq_id 1575197
- Location of start within SEQ ID NO 3023: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1406
- gi No. 5731259
- Description: (AF166351) alanine:glyoxylate aminotransferase 2 homolog [Arabidopsis thaliana]
- % Identity: 77
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 3025: from 1 to 71

Maximum Length Sequence:
related to:
Clone IDs:
222309
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3026
- Ceres seq_id 1575202
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3027
- Ceres seq_id 1575203
- Location of start within SEQ ID NO 3026: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3028
- Ceres seq_id 1575204
- Location of start within SEQ ID NO 3026: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 3028: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3029
- Ceres seq_id 1575205
- Location of start within SEQ ID NO 3026: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222353
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3030
- Ceres seq_id 1575210
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3031
- Ceres seq_id 1575211
- Location of start within SEQ ID NO 3030: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3032
- Ceres seq_id 1575212
- Location of start within SEQ ID NO 3030: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1407
- gi No. 2499335
- Description: NADH-UBIQUINONE OXIDOREDUCTASE 16 KD SUBUNIT (COMPLEX I-16KD) (CI-16KD)
- % Identity: 80
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3032: from 37 to 56

Maximum Length Sequence:
related to:
Clone IDs:
222418
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3033

- Ceres seq_id 1575217

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3034
- Ceres seq_id 1575218
- Location of start within SEQ ID NO 3033: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3034: from 20 to 93 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1408
- gi No. 688080
- Description: lectin=chitin-binding protein [Solanum tuberosum=potatoes, tubers, Peptide Partial, 27 aa]
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 3034: from 74 to 89

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3035
- Ceres seq_id 1575219
- Location of start within SEQ ID NO 3033: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222475

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3036
- Ceres seq_id 1575232

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3037
- Ceres seq_id 1575233
- Location of start within SEQ ID NO 3036: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1409
- gi No. 128388
- Description: NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (LTP) (PHOSPHOLIPID TRANSFER PROTEIN) (PLTP) >gi|82711|pir||A31779 phospholipid transfer protein 9C2 precursor - maize >gi|168576 (J04176) phospholipid transfer protein precursor [Zea mays]
- % Identity: 80
- Alignment Length: 25
- Location of Alignment in SEQ ID NO 3037: from 1 to 25

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3038
- Ceres seq_id 1575234
- Location of start within SEQ ID NO 3036: at 497 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222501
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3039
- Ceres seq_id 1575235
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3040
- Ceres seq_id 1575236
- Location of start within SEQ ID NO 3039: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family
- Location within SEQ ID NO 3040: from 93 to 182 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3041
- Ceres seq_id 1575237
- Location of start within SEQ ID NO 3039: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family
- Location within SEQ ID NO 3041: from 70 to 159 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3042
- Ceres seq_id 1575238
- Location of start within SEQ ID NO 3039: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Uncharacterized protein family
- Location within SEQ ID NO 3042: from 18 to 107 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
222544
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3043
- Ceres seq_id 1575243
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3044
- Ceres seq_id 1575244
- Location of start within SEQ ID NO 3043: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1410

- gi No. 400249
- Description: MFS14 PROTEIN PRECURSOR >gi|100873|pir||S25104 MSF14 protein - maize >gi|22645|emb|CAA47737| (X67323) MSF14 [Zea mays]
- % Identity: 94.4
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 3044: from 17 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3045
- Ceres seq_id 1575245
- Location of start within SEQ ID NO 3043: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3046
- Ceres seq_id 1575246
- Location of start within SEQ ID NO 3043: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1411
- gi No. 400249
- Description: MFS14 PROTEIN PRECURSOR >gi|100873|pir||S25104 MSF14 protein - maize >gi|22645|emb|CAA47737| (X67323) MSF14 [Zea mays]
- % Identity: 94.4
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 3046: from 1 to 36

Maximum Length Sequence:
related to:
Clone IDs:
223560
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3047
- Ceres seq_id 1575336

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3048
- Ceres seq_id 1575337
- Location of start within SEQ ID NO 3047: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Single-strand binding protein family
- Location within SEQ ID NO 3048: from 151 to 247 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1412
- gi No. 3600051
- Description: (AF080120) contains similarity to the single-strand binding proteins family (Pfam: SSB.hmm, score: 24.02) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 40
- Location of Alignment in SEQ ID NO 3048: from 187 to 226

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3049
- Ceres seq_id 1575338

- Location of start within SEQ ID NO 3047: at 234 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Single-strand binding protein family
- Location within SEQ ID NO 3049: from 74 to 170 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1413
- gi No. 3600051
- Description: (AF080120) contains similarity to the single-strand binding proteins family (Pfam: SSB.hmm, score: 24.02) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 40
- Location of Alignment in SEQ ID NO 3049: from 110 to 149

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3050
- Ceres seq_id 1575339
- Location of start within SEQ ID NO 3047: at 279 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Single-strand binding protein family
- Location within SEQ ID NO 3050: from 59 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1414
- gi No. 3600051
- Description: (AF080120) contains similarity to the single-strand binding proteins family (Pfam: SSB.hmm, score: 24.02) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 40
- Location of Alignment in SEQ ID NO 3050: from 95 to 134

Maximum Length Sequence:
related to:
Clone IDs:
223591
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3051
- Ceres seq_id 1575343

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3052
- Ceres seq_id 1575344
- Location of start within SEQ ID NO 3051: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1415
- gi No. 399783
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN (CLONE W10-1) >gi|388257|emb|CAA39222| (X55693) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 84.6
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3052: from 167 to 179

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3053

- Ceres seq_id 1575345
- Location of start within SEQ ID NO 3051: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1416
- gi No. 399783
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN (CLONE W10-1) >gi|388257|emb|CAA39222| (X55693) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 84.6
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3053: from 132 to 144

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3054
- Ceres seq_id 1575346
- Location of start within SEQ ID NO 3051: at 172 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1417
- gi No. 399783
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN (CLONE W10-1) >gi|388257|emb|CAA39222| (X55693) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 84.6
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3054: from 110 to 122

Maximum Length Sequence:
related to:
Clone IDs:
223645
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3055
- Ceres seq_id 1575351
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3056
- Ceres seq_id 1575352
- Location of start within SEQ ID NO 3055: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3057
- Ceres seq_id 1575353
- Location of start within SEQ ID NO 3055: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3057: from 44 to 131 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3058
- Ceres seq_id 1575354
- Location of start within SEQ ID NO 3055: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
223731
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3059
- Ceres seq_id 1575361
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3060
- Ceres seq_id 1575362
- Location of start within SEQ ID NO 3059: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 3060: from 68 to 125 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3061
- Ceres seq_id 1575363
- Location of start within SEQ ID NO 3059: at 5 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 3061: from 67 to 124 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3062
- Ceres seq_id 1575364
- Location of start within SEQ ID NO 3059: at 167 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 3062: from 13 to 70 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
224062
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3063
- Ceres seq_id 1575386
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3064
- Ceres seq_id 1575387

- Location of start within SEQ ID NO 3063: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1418
- gi No. 119714
- Description: EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN) >gi|82169|pir||S06733 hydroxyproline-rich glycoprotein precursor - common tobacco >gi|19867|emb|CAA32090| (X13885) extensin (AA 1-620) [Nicotiana tabacum]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3064: from 176 to 187

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3065
- Ceres seq_id 1575388
- Location of start within SEQ ID NO 3063: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3066
- Ceres seq_id 1575389
- Location of start within SEQ ID NO 3063: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
224210
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3067
- Ceres seq_id 1575402

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3068
- Ceres seq_id 1575403
- Location of start within SEQ ID NO 3067: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3068: from 93 to 135 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
224247
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3069
- Ceres seq_id 1575404

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3070

- Ceres seq_id 1575405
- Location of start within SEQ ID NO 3069: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oxysterol-binding protein
- Location within SEQ ID NO 3070: from 19 to 124 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3071
- Ceres seq_id 1575406
- Location of start within SEQ ID NO 3069: at 213 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3072
- Ceres seq_id 1575407
- Location of start within SEQ ID NO 3069: at 234 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
224565
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3073
- Ceres seq_id 1575428
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3074
- Ceres seq_id 1575429
- Location of start within SEQ ID NO 3073: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1419
- gi No. 4732091
- Description: (AF126742) bundle sheath defective protein 2 [Zea mays]
- % Identity: 86.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3074: from 21 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3075
- Ceres seq_id 1575430
- Location of start within SEQ ID NO 3073: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1420
- gi No. 4732091

- Description: (AF126742) bundle sheath defective protein 2 [Zea mays]
- % Identity: 86.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3075: from 1 to 114

Maximum Length Sequence:
related to:
Clone IDs:
224592
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3076
- Ceres seq_id 1575435
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3077
- Ceres seq_id 1575436
- Location of start within SEQ ID NO 3076: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1421
- gi No. 3153821
- Description: (AF062655) plenty-of-prolines-101; POP101; SH3-philo-protein [Mus musculus]
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3077: from 49 to 68

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3078
- Ceres seq_id 1575437
- Location of start within SEQ ID NO 3076: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3079
- Ceres seq_id 1575438
- Location of start within SEQ ID NO 3076: at 225 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
224596
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3080
- Ceres seq_id 1575439
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3081
- Ceres seq_id 1575440
- Location of start within SEQ ID NO 3080: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- FKBP-type peptidyl-prolyl cis-trans isomerases
- Location within SEQ ID NO 3081: from 78 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1422
- gi No. 4758380
- Description: ref|NP_004107.1|pFKBP1B| FK506-binding protein 1B (12.6 kD) >gi|627561|pir||JC2188 peptidylprolyl isomerase (EC 5.2.1.8) FKBP1L, form 1 - human >gi|546582|bbs|147143 (S69815) peptidyl-prolyl cis-trans isomerase, PPIase=12 kda FK506 binding
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3081: from 195 to 205

Maximum Length Sequence:
related to:
Clone IDs:
224763
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3082
- Ceres seq_id 1575441
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3083
- Ceres seq_id 1575442
- Location of start within SEQ ID NO 3082: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3084
- Ceres seq_id 1575443
- Location of start within SEQ ID NO 3082: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3085
- Ceres seq_id 1575444
- Location of start within SEQ ID NO 3082: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Elongation factor Tu family
- Location within SEQ ID NO 3085: from 38 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1423
- gi No. 3924612
- Description: (AF069442) mitochondrial elongation factor Tu [Arabidopsis thaliana] >gi|4263511|gb|AAD15337| (AC004044) mitochondrial elongation factor Tu [Arabidopsis thaliana]
- % Identity: 92.5
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 3085: from 22 to 126

Maximum Length Sequence:
related to:

Clone IDs:
225086
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3086
- Ceres seq_id 1575460
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3087
- Ceres seq_id 1575461
- Location of start within SEQ ID NO 3086: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1424
- gi No. 1079136
- Description: RNA-binding protein cabeza - fruit fly (Drosophila melanogaster) >gi|532788 (U13178) RNA binding protein [Drosophila melanogaster] >gi|567106 (L37083) RNA binding protein [Drosophila melanogaster]
- % Identity: 90.9
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3087: from 60 to 70

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3088
- Ceres seq_id 1575462
- Location of start within SEQ ID NO 3086: at 131 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1425
- gi No. 1079136
- Description: RNA-binding protein cabeza - fruit fly (Drosophila melanogaster) >gi|532788 (U13178) RNA binding protein [Drosophila melanogaster] >gi|567106 (L37083) RNA binding protein [Drosophila melanogaster]
- % Identity: 90.9
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3088: from 17 to 27

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3089
- Ceres seq_id 1575463
- Location of start within SEQ ID NO 3086: at 164 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1426
- gi No. 1079136
- Description: RNA-binding protein cabeza - fruit fly (Drosophila melanogaster) >gi|532788 (U13178) RNA binding protein [Drosophila melanogaster] >gi|567106 (L37083) RNA binding protein [Drosophila melanogaster]
- % Identity: 90.9
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3089: from 6 to 16

Maximum Length Sequence:
related to:

Clone IDs:
225274
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3090
- Ceres seq_id 1575488
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3091
- Ceres seq_id 1575489
- Location of start within SEQ ID NO 3090: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3092
- Ceres seq_id 1575490
- Location of start within SEQ ID NO 3090: at 15 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3093
- Ceres seq_id 1575491
- Location of start within SEQ ID NO 3090: at 265 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Germin family
- Location within SEQ ID NO 3093: from 1 to 132 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
225287
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3094
- Ceres seq_id 1575492
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3095
- Ceres seq_id 1575493
- Location of start within SEQ ID NO 3094: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3095: from 245 to 284 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3096
- Ceres seq_id 1575494
- Location of start within SEQ ID NO 3094: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3096: from 212 to 251 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3097
- Ceres seq_id 1575495
- Location of start within SEQ ID NO 3094: at 151 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3097: from 195 to 234 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
225321
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3098
- Ceres seq_id 1575496
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3099
- Ceres seq_id 1575497
- Location of start within SEQ ID NO 3098: at 187 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1427
- gi No. 2826882
- Description: (AJ223634) transcription factor IIA small subunit [Arabidopsis thaliana] >gi|5051786|emb|CAB45079.1| (AL078637) transcription factor IIA small subunit [Arabidopsis thaliana]
- % Identity: 82.1
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 3099: from 1 to 106

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3100
- Ceres seq_id 1575498
- Location of start within SEQ ID NO 3098: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1428
- gi No. 2826882
- Description: (AJ223634) transcription factor IIA small subunit [Arabidopsis thaliana] >gi|5051786|emb|CAB45079.1| (AL078637) transcription factor IIA small subunit [Arabidopsis thaliana]
- % Identity: 82.1
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 3100: from 1 to 93

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3101
- Ceres seq_id 1575499

- Location of start within SEQ ID NO 3098: at 253 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1429
- gi No. 2826882
- Description: (AJ223634) transcription factor IIA small subunit [Arabidopsis thaliana] >gi|5051786|emb|CAB45079.1| (AL078637) transcription factor IIA small subunit [Arabidopsis thaliana]
- % Identity: 82.1
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 3101: from 1 to 84

Maximum Length Sequence:
related to:
Clone IDs:
226282
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3102
- Ceres seq_id 1575556
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3103
- Ceres seq_id 1575557
- Location of start within SEQ ID NO 3102: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3104
- Ceres seq_id 1575558
- Location of start within SEQ ID NO 3102: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphopantetheine attachment site
- Location within SEQ ID NO 3104: from 56 to 110 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3105
- Ceres seq_id 1575559
- Location of start within SEQ ID NO 3102: at 239 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphopantetheine attachment site
- Location within SEQ ID NO 3105: from 18 to 72 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
226499
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3106
- Ceres seq_id 1575573

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3107
- Ceres seq_id 1575574
- Location of start within SEQ ID NO 3106: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3108
- Ceres seq_id 1575575
- Location of start within SEQ ID NO 3106: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3109
- Ceres seq_id 1575576
- Location of start within SEQ ID NO 3106: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1430
- gi No. 2134211
- Description: protamine II-5 - painted turtle
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3109: from 42 to 55

Maximum Length Sequence:
related to:
Clone IDs:
226564
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3110
- Ceres seq_id 1575588
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3111
- Ceres seq_id 1575589
- Location of start within SEQ ID NO 3110: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1431
- gi No. 3168840
- Description: (U88711) copper homeostasis factor [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 3111: from 79 to 168

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3112
- Ceres seq_id 1575590
- Location of start within SEQ ID NO 3110: at 224 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1432
- gi No. 3168840
- Description: (U88711) copper homeostasis factor [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 3112: from 5 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3113
- Ceres seq_id 1575591
- Location of start within SEQ ID NO 3110: at 260 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1433
- gi No. 3168840
- Description: (U88711) copper homeostasis factor [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 91
- Location of Alignment in SEQ ID NO 3113: from 1 to 82

Maximum Length Sequence:
related to:
Clone IDs:
226602
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3114
- Ceres seq_id 1575592
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3115
- Ceres seq_id 1575593
- Location of start within SEQ ID NO 3114: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1434
- gi No. 4769004
- Description: (AF140598) ring-box protein 1 [Homo sapiens] >gi|4769006|gb|AAD29716.1|AF140599_1 (AF140599) ring-box protein 1 [Mus musculus] >gi|4809216|gb|AAD30146.1|AF142059_1 (AF142059) RING finger protein [Homo sapiens]
- % Identity: 90.9
- Alignment Length: 99
- Location of Alignment in SEQ ID NO 3115: from 47 to 145

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3116
- Ceres seq_id 1575594
- Location of start within SEQ ID NO 3114: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3117
- Ceres seq_id 1575595
- Location of start within SEQ ID NO 3114: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1435
- gi No. 4769004
- Description: (AF140598) ring-box protein 1 [Homo sapiens] >gi|4769006|gb|AAD29716.1|AF140599_1 (AF140599) ring-box protein 1 [Mus musculus] >gi|4809216|gb|AAD30146.1|AF142059_1 (AF142059) RING finger protein [Homo sapiens]
- % Identity: 90.9
- Alignment Length: 99
- Location of Alignment in SEQ ID NO 3117: from 24 to 122

Maximum Length Sequence:
related to:
Clone IDs:
226620
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3118
- Ceres seq_id 1575596
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3119
- Ceres seq_id 1575597
- Location of start within SEQ ID NO 3118: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1436
- gi No. 2145052
- Description: (AF000412) merozoite surface protein-1 [Plasmodium berghei]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3119: from 47 to 59

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3120
- Ceres seq_id 1575598
- Location of start within SEQ ID NO 3118: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1437
- gi No. 2145052
- Description: (AF000412) merozoite surface protein-1 [Plasmodium berghei]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3120: from 28 to 40

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3121
- Ceres seq_id 1575599
- Location of start within SEQ ID NO 3118: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
226833
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3122
- Ceres seq_id 1575606
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3123
- Ceres seq_id 1575607
- Location of start within SEQ ID NO 3122: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 3123: from 119 to 195 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1438
- gi No. 1710151
- Description: (U72711) proline iminopeptidase [Arabidopsis thaliana]
- % Identity: 78.9
- Alignment Length: 147
- Location of Alignment in SEQ ID NO 3123: from 50 to 195

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3124
- Ceres seq_id 1575608
- Location of start within SEQ ID NO 3122: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3125
- Ceres seq_id 1575609
- Location of start within SEQ ID NO 3122: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- alpha/beta hydrolase fold
- Location within SEQ ID NO 3125: from 65 to 141 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1439
- gi No. 1710151
- Description: (U72711) proline iminopeptidase [Arabidopsis thaliana]
- % Identity: 78.9
- Alignment Length: 147
- Location of Alignment in SEQ ID NO 3125: from 1 to 141

Maximum Length Sequence:
related to:

Clone IDs:
227073
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3126
- Ceres seq_id 1575635
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3127
- Ceres seq_id 1575636
- Location of start within SEQ ID NO 3126: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3128
- Ceres seq_id 1575637
- Location of start within SEQ ID NO 3126: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1440
- gi No. 730547
- Description: 60S RIBOSOMAL PROTEIN L27
- % Identity: 81.4
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 3128: from 24 to 82

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3129
- Ceres seq_id 1575638
- Location of start within SEQ ID NO 3126: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1441
- gi No. 730547
- Description: 60S RIBOSOMAL PROTEIN L27
- % Identity: 81.4
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 3129: from 1 to 59

Maximum Length Sequence:
related to:
Clone IDs:
227278
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3130
- Ceres seq_id 1575645
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3131
- Ceres seq_id 1575646
- Location of start within SEQ ID NO 3130: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1442
- gi No. 5305594

- Description: (AF102988) Ca2+-independent phospholipase A2 short isoform [Homo sapiens]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3131: from 88 to 98

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3132
- Ceres seq_id 1575647
- Location of start within SEQ ID NO 3130: at 289 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3133
- Ceres seq_id 1575648
- Location of start within SEQ ID NO 3130: at 310 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
227627
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3134
- Ceres seq_id 1575682
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3135
- Ceres seq_id 1575683
- Location of start within SEQ ID NO 3134: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1443
- gi No. 2500399
- Description: 40S RIBOSOMAL PROTEIN S3 >gi|1836060|bbs|179561 (S83098) ribosomal protein S3 [Ambystoma mexicanum=Mexican axolotls, embryos, Peptide, 253 aa] [Ambystoma mexicanum]
- % Identity: 73.5
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 3135: from 63 to 96

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3136
- Ceres seq_id 1575684
- Location of start within SEQ ID NO 3134: at 39 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1444
- gi No. 2500399
- Description: 40S RIBOSOMAL PROTEIN S3 >gi|1836060|bbs|179561 (S83098) ribosomal protein S3 [Ambystoma mexicanum=Mexican axolotls, embryos, Peptide, 253 aa] [Ambystoma mexicanum]

- % Identity: 73.5
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 3136: from 51 to 84

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3137
- Ceres seq_id 1575685
- Location of start within SEQ ID NO 3134: at 463 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
227940
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3138
- Ceres seq_id 1575704
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3139
- Ceres seq_id 1575705
- Location of start within SEQ ID NO 3138: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S16
- Location within SEQ ID NO 3139: from 58 to 115 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3140
- Ceres seq_id 1575706
- Location of start within SEQ ID NO 3138: at 149 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S16
- Location within SEQ ID NO 3140: from 9 to 66 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3141
- Ceres seq_id 1575707
- Location of start within SEQ ID NO 3138: at 209 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S16
- Location within SEQ ID NO 3141: from 1 to 46 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
228245
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3142
- Ceres seq_id 1575720

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3143
- Ceres seq_id 1575721
- Location of start within SEQ ID NO 3142: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3143: from 8 to 132 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1445
- gi No. 422032
- Description: profilin 2 - maize >gi|313140|emb|CAA51719| (X73280) profilin 2 [Zea mays]
- % Identity: 98.5
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3143: from 1 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3144
- Ceres seq_id 1575722
- Location of start within SEQ ID NO 3142: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3144: from 2 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1446
- gi No. 422032
- Description: profilin 2 - maize >gi|313140|emb|CAA51719| (X73280) profilin 2 [Zea mays]
- % Identity: 98.5
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3144: from 1 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3145
- Ceres seq_id 1575723
- Location of start within SEQ ID NO 3142: at 137 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3145: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1447
- gi No. 422032
- Description: profilin 2 - maize >gi|313140|emb|CAA51719| (X73280) profilin 2 [Zea mays]
- % Identity: 98.5
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3145: from 1 to 120

Maximum Length Sequence:
related to:

Clone IDs:
228406
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3146
- Ceres seq_id 1575729
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3147
- Ceres seq_id 1575730
- Location of start within SEQ ID NO 3146: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Skp1 family
- Location within SEQ ID NO 3147: from 46 to 113 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3148
- Ceres seq_id 1575731
- Location of start within SEQ ID NO 3146: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3149
- Ceres seq_id 1575732
- Location of start within SEQ ID NO 3146: at 82 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Skp1 family
- Location within SEQ ID NO 3149: from 19 to 86 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
228700
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3150
- Ceres seq_id 1575740
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3151
- Ceres seq_id 1575741
- Location of start within SEQ ID NO 3150: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3152
- Ceres seq_id 1575742
- Location of start within SEQ ID NO 3150: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1448
- gi No. 1185556
- Description: (U45859) glyceraldehyde-3-phosphate dehydrogenase [Zea mays]
- % Identity: 70.4
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3152: from 1 to 27

Maximum Length Sequence:
related to:
Clone IDs:
228879
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3153
- Ceres seq_id 1575750
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3154
- Ceres seq_id 1575751
- Location of start within SEQ ID NO 3153: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SCP-like extracellular protein
- Location within SEQ ID NO 3154: from 4 to 142 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1449
- gi No. 130940
- Description: PATHOGENESIS-RELATED PROTEIN PRMS PRECURSOR
>gi|100906|pir||S14969 pathogenesis-related protein - maize
>gi|22454|emb|CAA38223| (X54325) pathogenesis-related protein [Zea mays]
- % Identity: 90.2
- Alignment Length: 144
- Location of Alignment in SEQ ID NO 3154: from 1 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3155
- Ceres seq_id 1575752
- Location of start within SEQ ID NO 3153: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SCP-like extracellular protein
- Location within SEQ ID NO 3155: from 1 to 127 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1450
- gi No. 130940
- Description: PATHOGENESIS-RELATED PROTEIN PRMS PRECURSOR
>gi|100906|pir||S14969 pathogenesis-related protein - maize
>gi|22454|emb|CAA38223| (X54325) pathogenesis-related protein [Zea mays]
- % Identity: 90.2
- Alignment Length: 144
- Location of Alignment in SEQ ID NO 3155: from 1 to 127

Maximum Length Sequence:
related to:
Clone IDs:
230028
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3156
- Ceres seq_id 1575794

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3157
- Ceres seq_id 1575795
- Location of start within SEQ ID NO 3156: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3157: from 41 to 172 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1451
- gi No. 1362107
- Description: GUT8-2a protein - common tobacco
- % Identity: 90.9
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 3157: from 44 to 87

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3158
- Ceres seq_id 1575796
- Location of start within SEQ ID NO 3156: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3158: from 10 to 141 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1452
- gi No. 1362107
- Description: GUT8-2a protein - common tobacco
- % Identity: 90.9
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 3158: from 13 to 56

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3159
- Ceres seq_id 1575797
- Location of start within SEQ ID NO 3156: at 164 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3159: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1453
- gi No. 1362107
- Description: GUT8-2a protein - common tobacco
- % Identity: 90.9
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 3159: from 1 to 33

Maximum Length Sequence:
related to:
Clone IDs:
230239
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3160
- Ceres seq_id 1575798

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3161
- Ceres seq_id 1575799
- Location of start within SEQ ID NO 3160: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3162
- Ceres seq_id 1575800
- Location of start within SEQ ID NO 3160: at 383 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1454
- gi No. 3041724
- Description: 26S PROTEASE REGULATORY SUBUNIT 8 (SUG1 HOMOLOG) (XSUG1) >gi|1877414|emb|CAA57512| (X81986) XSUG1 [Xenopus laevis]
- % Identity: 76
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 3162: from 1 to 107

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3163
- Ceres seq_id 1575801
- Location of start within SEQ ID NO 3160: at 449 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1455
- gi No. 3041724
- Description: 26S PROTEASE REGULATORY SUBUNIT 8 (SUG1 HOMOLOG) (XSUG1) >gi|1877414|emb|CAA57512| (X81986) XSUG1 [Xenopus laevis]
- % Identity: 76
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 3163: from 1 to 85

Maximum Length Sequence:
related to:
Clone IDs:
230701

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3164
- Ceres seq_id 1575815

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3165
- Ceres seq_id 1575816
- Location of start within SEQ ID NO 3164: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3165: from 28 to 128 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3166
- Ceres seq_id 1575817
- Location of start within SEQ ID NO 3164: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3166: from 5 to 105 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3167
- Ceres seq_id 1575818
- Location of start within SEQ ID NO 3164: at 234 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3167: from 1 to 51 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
230717
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3168
- Ceres seq_id 1575822

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3169
- Ceres seq_id 1575823
- Location of start within SEQ ID NO 3168: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin family
- Location within SEQ ID NO 3169: from 1 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1456
- gi No. 49586
- Description: (X08013) ubiquitin (752 is 2nd base in codon) [Cricetulus sp.]
- % Identity: 70
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 3169: from 1 to 70

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3170
- Ceres seq_id 1575824
- Location of start within SEQ ID NO 3168: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1457
- gi No. 1841849

- Description: (U86866) polyubiquitin [Actinidia chinensis]
- % Identity: 70.7
- Alignment Length: 41
- Location of Alignment in SEQ ID NO 3170: from 1 to 11

Maximum Length Sequence:
related to:
Clone IDs:
230729
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3171
- Ceres seq_id 1575825

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3172
- Ceres seq_id 1575826
- Location of start within SEQ ID NO 3171: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 3172: from 31 to 91 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3173
- Ceres seq_id 1575827
- Location of start within SEQ ID NO 3171: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 3173: from 23 to 83 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3174
- Ceres seq_id 1575828
- Location of start within SEQ ID NO 3171: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S10
- Location within SEQ ID NO 3174: from 19 to 79 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
232679
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3175
- Ceres seq_id 1575844

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3176
- Ceres seq_id 1575845
- Location of start within SEQ ID NO 3175: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3176: from 97 to 137 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3177
- Ceres seq_id 1575846
- Location of start within SEQ ID NO 3175: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3177: from 92 to 132 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3178
- Ceres seq_id 1575847
- Location of start within SEQ ID NO 3175: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3178: from 48 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
232797
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3179
- Ceres seq_id 1575856
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3180
- Ceres seq_id 1575857
- Location of start within SEQ ID NO 3179: at 330 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1458
- gi No. 462338
- Description: SPERM PROTAMINE P1 >gi|1071970|pir||S34045 protamine - North American opossum >gi|294036 (L17007) protamine 1 [Didelphis marsupialis] >gi|407063|emb|CAA52193| (X74044) protamine P1 [Didelphis marsupialis] >gi|598253 (L35448) protamine P1
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3180: from 70 to 89

Maximum Length Sequence:
related to:
Clone IDs:
233659
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3181
- Ceres seq_id 1575935

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3182
- Ceres seq_id 1575936
- Location of start within SEQ ID NO 3181: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 3182: from 1 to 233 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1459
- gi No. 2191149
- Description: (AF007269) Similar to protein kinase [Arabidopsis thaliana]
- % Identity: 81
- Alignment Length: 242
- Location of Alignment in SEQ ID NO 3182: from 1 to 242

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3183
- Ceres seq_id 1575937
- Location of start within SEQ ID NO 3181: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 3183: from 1 to 202 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1460
- gi No. 2191149
- Description: (AF007269) Similar to protein kinase [Arabidopsis thaliana]
- % Identity: 81
- Alignment Length: 242
- Location of Alignment in SEQ ID NO 3183: from 1 to 211

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3184
- Ceres seq_id 1575938
- Location of start within SEQ ID NO 3181: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 3184: from 1 to 171 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1461
- gi No. 2191149
- Description: (AF007269) Similar to protein kinase [Arabidopsis thaliana]
- % Identity: 81
- Alignment Length: 242
- Location of Alignment in SEQ ID NO 3184: from 1 to 180

Maximum Length Sequence:
related to:

Clone IDs:
233992
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3185
- Ceres seq_id 1575954
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3186
- Ceres seq_id 1575955
- Location of start within SEQ ID NO 3185: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Small hydrophilic plant seed proteins
- Location within SEQ ID NO 3186: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1462
- gi No. 168467
- Description: (M90554) late embryogenesis abundant protein [Zea mays]>gi|445606|prf||1909353A Emb5 gene [Zea mays]
- % Identity: 94.7
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 3186: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3187
- Ceres seq_id 1575956
- Location of start within SEQ ID NO 3185: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Small hydrophilic plant seed proteins
- Location within SEQ ID NO 3187: from 1 to 68 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1463
- gi No. 168467
- Description: (M90554) late embryogenesis abundant protein [Zea mays]>gi|445606|prf||1909353A Emb5 gene [Zea mays]
- % Identity: 94.7
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 3187: from 1 to 100

Maximum Length Sequence:
related to:
Clone IDs:
234181
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3188
- Ceres seq_id 1575963
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3189
- Ceres seq_id 1575964
- Location of start within SEQ ID NO 3188: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ central domain (4 repeats)
- Location within SEQ ID NO 3189: from 84 to 143 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3190
- Ceres seq_id 1575965
- Location of start within SEQ ID NO 3188: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ central domain (4 repeats)
- Location within SEQ ID NO 3190: from 63 to 122 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3191
- Ceres seq_id 1575966
- Location of start within SEQ ID NO 3188: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ central domain (4 repeats)
- Location within SEQ ID NO 3191: from 61 to 120 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234447
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3192
- Ceres seq_id 1575992

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3193
- Ceres seq_id 1575993
- Location of start within SEQ ID NO 3192: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3194
- Ceres seq_id 1575994
- Location of start within SEQ ID NO 3192: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3195
- Ceres seq_id 1575995
- Location of start within SEQ ID NO 3192: at 532 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Retroviral aspartyl proteases
- Location within SEQ ID NO 3195: from 1 to 55 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234604
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3196
- Ceres seq_id 1576002
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3197
- Ceres seq_id 1576003
- Location of start within SEQ ID NO 3196: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L19
- Location within SEQ ID NO 3197: from 136 to 236 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3198
- Ceres seq_id 1576004
- Location of start within SEQ ID NO 3196: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L19
- Location within SEQ ID NO 3198: from 106 to 206 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234758
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3199
- Ceres seq_id 1576026
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3200
- Ceres seq_id 1576027
- Location of start within SEQ ID NO 3199: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L23
- Location within SEQ ID NO 3200: from 138 to 217 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1464
- gi No. 585876
- Description: 60S RIBOSOMAL PROTEIN L23A (L25) >gi|1084424|pir||S48026 ribosomal protein L25 - common tobacco >gi|310935 (L18908) 60S ribosomal protein L25 [Nicotiana tabacum]
- % Identity: 83.8
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 3200: from 70 to 220

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3201

- Ceres seq_id 1576028
- Location of start within SEQ ID NO 3199: at 208 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L23
- Location within SEQ ID NO 3201: from 69 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1465
- gi No. 585876
- Description: 60S RIBOSOMAL PROTEIN L23A (L25) >gi|1084424|pir||S48026 ribosomal protein L25 - common tobacco >gi|310935 (L18908) 60S ribosomal protein L25 [Nicotiana tabacum]
- % Identity: 83.8
- Alignment Length: 154
- Location of Alignment in SEQ ID NO 3201: from 1 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3202
- Ceres seq_id 1576029
- Location of start within SEQ ID NO 3199: at 623 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234879
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3203
- Ceres seq_id 1576041

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3204
- Ceres seq_id 1576042
- Location of start within SEQ ID NO 3203: at 210 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3205
- Ceres seq_id 1576043
- Location of start within SEQ ID NO 3203: at 361 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3206
- Ceres seq_id 1576044
- Location of start within SEQ ID NO 3203: at 464 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3206: from 45 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234899
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3207
- Ceres seq_id 1576045
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3208
- Ceres seq_id 1576046
- Location of start within SEQ ID NO 3207: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 3208: from 27 to 130 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3209
- Ceres seq_id 1576047
- Location of start within SEQ ID NO 3207: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3210
- Ceres seq_id 1576048
- Location of start within SEQ ID NO 3207: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 3210: from 1 to 104 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234912
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3211
- Ceres seq_id 1576049
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3212
- Ceres seq_id 1576050
- Location of start within SEQ ID NO 3211: at 354 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 3212: from 24 to 90 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3213
- Ceres seq_id 1576051
- Location of start within SEQ ID NO 3211: at 414 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 3213: from 4 to 70 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3214
- Ceres seq_id 1576052
- Location of start within SEQ ID NO 3211: at 417 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Clathrin adaptor complex small chain
- Location within SEQ ID NO 3214: from 3 to 69 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
234987
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3215
- Ceres seq_id 1576067

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3216
- Ceres seq_id 1576068
- Location of start within SEQ ID NO 3215: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3217
- Ceres seq_id 1576069
- Location of start within SEQ ID NO 3215: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3218
- Ceres seq_id 1576070
- Location of start within SEQ ID NO 3215: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1466
- gi No. 3717946
- Description: (AJ005901) vag1 [Arabidopsis thaliana]

- % Identity: 73.6
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 3218: from 1 to 110

Maximum Length Sequence:
related to:
Clone IDs:
235129
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3219
- Ceres seq_id 1576075
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3220
- Ceres seq_id 1576076
- Location of start within SEQ ID NO 3219: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L5
- Location within SEQ ID NO 3220: from 33 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1467
- gi No. 1173055
- Description: 60S RIBOSOMAL PROTEIN L11 (L5)
>gi|541961|pir||S42497 ribosomal protein L11.e - alfalfa
>gi|1076504|pir||S51819 RL5 ribosomal protein - alfalfa
>gi|463252|emb|CAA55090| (X78284) RL5 ribosomal protein [Medicago sativa]
- % Identity: 74.3
- Alignment Length: 101
- Location of Alignment in SEQ ID NO 3220: from 25 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3221
- Ceres seq_id 1576077
- Location of start within SEQ ID NO 3219: at 322 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3222
- Ceres seq_id 1576078
- Location of start within SEQ ID NO 3219: at 325 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
235353
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3223
- Ceres seq_id 1576102
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3224
- Ceres seq_id 1576103
- Location of start within SEQ ID NO 3223: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3225
- Ceres seq_id 1576104
- Location of start within SEQ ID NO 3223: at 205 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1468
- gi No. 1922964
- Description: (AC000106) Similar to Schizosaccharomyces CCAAT-binding factor (gb|U88525). EST gb|T04310 comes from this gene. [Arabidopsis thaliana]
- % Identity: 81.6
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 3225: from 65 to 139

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3226
- Ceres seq_id 1576105
- Location of start within SEQ ID NO 3223: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1469
- gi No. 1922964
- Description: (AC000106) Similar to Schizosaccharomyces CCAAT-binding factor (gb|U88525). EST gb|T04310 comes from this gene. [Arabidopsis thaliana]
- % Identity: 81.6
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 3226: from 58 to 132

Maximum Length Sequence:
related to:
Clone IDs:
235378

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3227
- Ceres seq_id 1576110

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3228
- Ceres seq_id 1576111
- Location of start within SEQ ID NO 3227: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3228: from 91 to 179 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3229
- Ceres seq_id 1576112

- Location of start within SEQ ID NO 3227: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3229: from 54 to 142 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
235706
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3230
- Ceres seq_id 1576129
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3231
- Ceres seq_id 1576130
- Location of start within SEQ ID NO 3230: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 3231: from 138 to 177 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3232
- Ceres seq_id 1576131
- Location of start within SEQ ID NO 3230: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 3232: from 81 to 120 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
235730
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3233
- Ceres seq_id 1576134
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3234
- Ceres seq_id 1576135
- Location of start within SEQ ID NO 3233: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 3234: from 1 to 96 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3235

- Ceres seq_id 1576137
- Location of start within SEQ ID NO 3233: at 15 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc-binding dehydrogenases
- Location within SEQ ID NO 3235: from 1 to 92 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
235744
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3236
- Ceres seq_id 1576138
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3237
- Ceres seq_id 1576139
- Location of start within SEQ ID NO 3236: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1470
- gi No. 3122724
- Description: 60S RIBOSOMAL PROTEIN L38 >gi|2289009 (AC002335) ribosomal protein L38 isolog [Arabidopsis thaliana]
- % Identity: 85.5
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 3237: from 32 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3238
- Ceres seq_id 1576140
- Location of start within SEQ ID NO 3236: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1471
- gi No. 3122724
- Description: 60S RIBOSOMAL PROTEIN L38 >gi|2289009 (AC002335) ribosomal protein L38 isolog [Arabidopsis thaliana]
- % Identity: 85.5
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 3238: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
235776
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3239
- Ceres seq_id 1576144
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3240
- Ceres seq_id 1576145
- Location of start within SEQ ID NO 3239: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 3240: from 40 to 246 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1472
- gi No. 2645170
- Description: (D50576) YY2 protein [Oryza sativa]
- % Identity: 83.4
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 3240: from 36 to 246

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3241
- Ceres seq_id 1576146
- Location of start within SEQ ID NO 3239: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 3241: from 7 to 213 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1473
- gi No. 2645170
- Description: (D50576) YY2 protein [Oryza sativa]
- % Identity: 83.4
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 3241: from 3 to 213

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3242
- Ceres seq_id 1576147
- Location of start within SEQ ID NO 3239: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chalcone and stilbene synthases
- Location within SEQ ID NO 3242: from 1 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1474
- gi No. 2645170
- Description: (D50576) YY2 protein [Oryza sativa]
- % Identity: 83.4
- Alignment Length: 211
- Location of Alignment in SEQ ID NO 3242: from 1 to 193

Maximum Length Sequence:
related to:
Clone IDs:
235881
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3243
- Ceres seq_id 1576155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3244
- Ceres seq_id 1576156
- Location of start within SEQ ID NO 3243: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Bacterial transferase hexapeptide (four repeats)
- Location within SEQ ID NO 3244: from 43 to 76 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3245
- Ceres seq_id 1576157
- Location of start within SEQ ID NO 3243: at 596 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3246
- Ceres seq_id 1576158
- Location of start within SEQ ID NO 3243: at 656 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
235891
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3247
- Ceres seq_id 1576159
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3248
- Ceres seq_id 1576160
- Location of start within SEQ ID NO 3247: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1475
- gi No. 81869
- Description: hydroxyproline-rich glycoprotein (clone Hyp2.13) - kidney bean (fragment) >gi|169349 (M18095) hydroxyproline-rich glycoprotein [Phaseolus vulgaris]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3248: from 82 to 92

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3249
- Ceres seq_id 1576161
- Location of start within SEQ ID NO 3247: at 46 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1476
- gi No. 81869

- Description: hydroxyproline-rich glycoprotein (clone Hyp2.13) - kidney bean (fragment) >gi|169349 (M18095) hydroxyproline-rich glycoprotein [Phaseolus vulgaris]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3249: from 67 to 77

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3250
- Ceres seq_id 1576162
- Location of start within SEQ ID NO 3247: at 307 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
236569
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3251
- Ceres seq_id 1576221
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3252
- Ceres seq_id 1576222
- Location of start within SEQ ID NO 3251: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG14 and HMG17
- Location within SEQ ID NO 3252: from 42 to 147 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3253
- Ceres seq_id 1576223
- Location of start within SEQ ID NO 3251: at 235 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG14 and HMG17
- Location within SEQ ID NO 3253: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3254
- Ceres seq_id 1576224
- Location of start within SEQ ID NO 3251: at 260 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
236621
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3255
- Ceres seq_id 1576229

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3256
- Ceres seq_id 1576230
- Location of start within SEQ ID NO 3255: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1477
- gi No. 22380
- Description: (X59714) CAAT-box DNA binding protein subunit B (NF-YB) [Zea mays]
- % Identity: 84.3
- Alignment Length: 51
- Location of Alignment in SEQ ID NO 3256: from 61 to 106

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3257
- Ceres seq_id 1576231
- Location of start within SEQ ID NO 3255: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1478
- gi No. 22380
- Description: (X59714) CAAT-box DNA binding protein subunit B (NF-YB) [Zea mays]
- % Identity: 84.3
- Alignment Length: 51
- Location of Alignment in SEQ ID NO 3257: from 5 to 50

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3258
- Ceres seq_id 1576232
- Location of start within SEQ ID NO 3255: at 572 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
236747

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3259
- Ceres seq_id 1576241

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3260
- Ceres seq_id 1576242
- Location of start within SEQ ID NO 3259: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pterin 4 alpha carbinolamine dehydratase
- Location within SEQ ID NO 3260: from 119 to 215 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3261
- Ceres seq_id 1576243
- Location of start within SEQ ID NO 3259: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pterin 4 alpha carbinolamine dehydratase
- Location within SEQ ID NO 3261: from 83 to 179 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3262
- Ceres seq_id 1576244
- Location of start within SEQ ID NO 3259: at 246 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pterin 4 alpha carbinolamine dehydratase
- Location within SEQ ID NO 3262: from 38 to 134 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
237036
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3263
- Ceres seq_id 1576245

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3264
- Ceres seq_id 1576246
- Location of start within SEQ ID NO 3263: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1479
- gi No. 2425066
- Description: (AF019147) cysteine proteinase Mir3 [Zea mays]
- % Identity: 93.8
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 3264: from 67 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3265
- Ceres seq_id 1576247
- Location of start within SEQ ID NO 3263: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3265: from 92 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1480
- gi No. 2425066
- Description: (AF019147) cysteine proteinase Mir3 [Zea mays]

- % Identity: 85
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3265: from 5 to 24

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3266
- Ceres seq_id 1576248
- Location of start within SEQ ID NO 3263: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1481
- gi No. 2425066
- Description: (AF019147) cysteine proteinase Mir3 [Zea mays]
- % Identity: 93.8
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 3266: from 1 to 65

Maximum Length Sequence:
related to:
Clone IDs:
237666

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3267
- Ceres seq_id 1576267

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3268
- Ceres seq_id 1576268
- Location of start within SEQ ID NO 3267: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 3268: from 1 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1482
- gi No. 3193292
- Description: (AF069298) similar to ATPases associated with various cellular activites (Pfam: AAA.hmm, score: 230.91) [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 157
- Location of Alignment in SEQ ID NO 3268: from 1 to 157

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3269
- Ceres seq_id 1576269
- Location of start within SEQ ID NO 3267: at 8 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 3269: from 1 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1483
- gi No. 3193292

- Description: (AF069298) similar to ATPases associated with various cellular activites (Pfam: AAA.hmm, score: 230.91) [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 157
- Location of Alignment in SEQ ID NO 3269: from 1 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3270
- Ceres seq_id 1576270
- Location of start within SEQ ID NO 3267: at 47 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 3270: from 1 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1484
- gi No. 3193292
- Description: (AF069298) similar to ATPases associated with various cellular activites (Pfam: AAA.hmm, score: 230.91) [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 157
- Location of Alignment in SEQ ID NO 3270: from 1 to 142

Maximum Length Sequence:
related to:
Clone IDs:
237742
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3271
- Ceres seq_id 1576273
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3272
- Ceres seq_id 1576274
- Location of start within SEQ ID NO 3271: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1485
- gi No. 2677830
- Description: (U93168) ribosomal protein L12 [Prunus armeniaca]
- % Identity: 89.6
- Alignment Length: 48
- Location of Alignment in SEQ ID NO 3272: from 170 to 217

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3273
- Ceres seq_id 1576275
- Location of start within SEQ ID NO 3271: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L11
- Location within SEQ ID NO 3273: from 65 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1486

- gi No. 2677830
- Description: (U93168) ribosomal protein L12 [Prunus armeniaca]
- % Identity: 95.1
- Alignment Length: 103
- Location of Alignment in SEQ ID NO 3273: from 53 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3274
- Ceres seq_id 1576276
- Location of start within SEQ ID NO 3271: at 159 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L11
- Location within SEQ ID NO 3274: from 13 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1487
- gi No. 2677830
- Description: (U93168) ribosomal protein L12 [Prunus armeniaca]
- % Identity: 95.1
- Alignment Length: 103
- Location of Alignment in SEQ ID NO 3274: from 1 to 103

Maximum Length Sequence:
related to:
Clone IDs:
238765
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3275
- Ceres seq_id 1576315

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3276
- Ceres seq_id 1576316
- Location of start within SEQ ID NO 3275: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1488
- gi No. 1644232
- Description: (D67066) N-WASP [Bos taurus]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3276: from 85 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3277
- Ceres seq_id 1576317
- Location of start within SEQ ID NO 3275: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3278
- Ceres seq_id 1576318
- Location of start within SEQ ID NO 3275: at 32 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1489
- gi No. 1644232
- Description: (D67066) N-WASP [Bos taurus]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3278: from 75 to 90

Maximum Length Sequence:
related to:
Clone IDs:
239046
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3279
- Ceres seq_id 1576329
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3280
- Ceres seq_id 1576330
- Location of start within SEQ ID NO 3279: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 3280: from 8 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1490
- gi No. 2267597
- Description: (AF009413) 10 kDa chaperonin [Oryza sativa]
- % Identity: 86.5
- Alignment Length: 96
- Location of Alignment in SEQ ID NO 3280: from 6 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3281
- Ceres seq_id 1576331
- Location of start within SEQ ID NO 3279: at 18 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chaperonins 10 Kd subunit
- Location within SEQ ID NO 3281: from 3 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1491
- gi No. 2267597
- Description: (AF009413) 10 kDa chaperonin [Oryza sativa]
- % Identity: 86.5
- Alignment Length: 96
- Location of Alignment in SEQ ID NO 3281: from 1 to 95

Maximum Length Sequence:
related to:
Clone IDs:
239188
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3282
- Ceres seq_id 1576334
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3283
- Ceres seq_id 1576335
- Location of start within SEQ ID NO 3282: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3283: from 20 to 128 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1492
- gi No. 1419611
- Description: (X98857) HMG [Lampetra fluviatilis]
- % Identity: 70.8
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 3283: from 96 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3284
- Ceres seq_id 1576336
- Location of start within SEQ ID NO 3282: at 335 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3285
- Ceres seq_id 1576337
- Location of start within SEQ ID NO 3282: at 353 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
239517
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3286
- Ceres seq_id 1576347
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3287
- Ceres seq_id 1576348
- Location of start within SEQ ID NO 3286: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 3287: from 40 to 120 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1493
- gi No. 2407800
- Description: (Y12575) histone H2A.F/Z [Arabidopsis thaliana]
- % Identity: 81.7
- Alignment Length: 82
- Location of Alignment in SEQ ID NO 3287: from 39 to 120

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3288
- Ceres seq_id 1576349
- Location of start within SEQ ID NO 3286: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 3288: from 2 to 82 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1494
- gi No. 2407800
- Description: (Y12575) histone H2A.F/Z [Arabidopsis thaliana]
- % Identity: 81.7
- Alignment Length: 82
- Location of Alignment in SEQ ID NO 3288: from 1 to 82

Maximum Length Sequence:
related to:
Clone IDs:
239624
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3289
- Ceres seq_id 1576358

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3290
- Ceres seq_id 1576359
- Location of start within SEQ ID NO 3289: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3290: from 46 to 86 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3291
- Ceres seq_id 1576360
- Location of start within SEQ ID NO 3289: at 24 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3291: from 39 to 79 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3292
- Ceres seq_id 1576361
- Location of start within SEQ ID NO 3289: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C3HC4 type (RING finger)
- Location within SEQ ID NO 3292: from 21 to 61 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

related to:
Clone IDs:
239696
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3293
- Ceres seq_id 1576362
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3294
- Ceres seq_id 1576363
- Location of start within SEQ ID NO 3293: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1495
- gi No. 399854
- Description: HISTONE H2B.2 >gi|283042|pir||S28049 histone H2B - maize >gi|22325|emb|CAA40565| (X57313) H2B histone [Zea mays]
- % Identity: 93.5
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 3294: from 27 to 57

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3295
- Ceres seq_id 1576364
- Location of start within SEQ ID NO 3293: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3296
- Ceres seq_id 1576365
- Location of start within SEQ ID NO 3293: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
239871
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3297
- Ceres seq_id 1576372
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3298
- Ceres seq_id 1576373
- Location of start within SEQ ID NO 3297: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S21e
- Location within SEQ ID NO 3298: from 81 to 158 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1496
- gi No. 2500497

- Description: 40S RIBOSOMAL PROTEIN S21
>gi|1419372|emb|CAA67225.1| (X98656) ribosomal protein S21 [Zea mays]
- % Identity: 98.7
- Alignment Length: 78
- Location of Alignment in SEQ ID NO 3298: from 81 to 158

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3299
- Ceres seq_id 1576374
- Location of start within SEQ ID NO 3297: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S21e
- Location within SEQ ID NO 3299: from 53 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1497
- gi No. 2500497
- Description: 40S RIBOSOMAL PROTEIN S21
>gi|1419372|emb|CAA67225.1| (X98656) ribosomal protein S21 [Zea mays]
- % Identity: 98.7
- Alignment Length: 78
- Location of Alignment in SEQ ID NO 3299: from 53 to 130

Maximum Length Sequence:
related to:
Clone IDs:
239892
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3300
- Ceres seq_id 1576379
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3301
- Ceres seq_id 1576380
- Location of start within SEQ ID NO 3300: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1498
- gi No. 5734633
- Description: (AP000391) Similar to Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (AL031018) [Oryza sativa]
- % Identity: 86.8
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 3301: from 1 to 76

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3302
- Ceres seq_id 1576381
- Location of start within SEQ ID NO 3300: at 294 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3303
- Ceres seq_id 1576382
- Location of start within SEQ ID NO 3300: at 318 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
240338
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3304
- Ceres seq_id 1576401
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3305
- Ceres seq_id 1576402
- Location of start within SEQ ID NO 3304: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1499
- gi No. 123541
- Description: 17.4 KD CLASS I HEAT SHOCK PROTEIN (HSP 17.4) >gi|81639|pir||JQ0351 heat shock protein 17 - Arabidopsis thaliana >gi|16340|emb|CAA35182| (X17293) heat shock protein (AA 1 - 156) [Arabidopsis thaliana]
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3305: from 138 to 150

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3306
- Ceres seq_id 1576403
- Location of start within SEQ ID NO 3304: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 3306: from 68 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1500
- gi No. 1122317
- Description: (X94193) heat shock protein 17.9 [Pennisetum glaucum]
- % Identity: 82.2
- Alignment Length: 73
- Location of Alignment in SEQ ID NO 3306: from 68 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3307
- Ceres seq_id 1576404
- Location of start within SEQ ID NO 3304: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

240357
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3308
- Ceres seq_id 1576405
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3309
- Ceres seq_id 1576406
- Location of start within SEQ ID NO 3308: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3310
- Ceres seq_id 1576407
- Location of start within SEQ ID NO 3308: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1501
- gi No. 3043714
- Description: (AB011167) KIAA0595 protein [Homo sapiens]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 3310: from 27 to 38

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3311
- Ceres seq_id 1576408
- Location of start within SEQ ID NO 3308: at 37 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
240938
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3312
- Ceres seq_id 1576429
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3313
- Ceres seq_id 1576430
- Location of start within SEQ ID NO 3312: at 248 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1502
- gi No. 4226151
- Description: (AF125462) similar GAR1 related proteins (small nucleolar protein required for pre-rRNA splicing) [Caenorhabditis elegans]
- % Identity: 83.3
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 3313: from 32 to 43

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3314
- Ceres seq_id 1576431
- Location of start within SEQ ID NO 3312: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3315
- Ceres seq_id 1576432
- Location of start within SEQ ID NO 3312: at 281 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1503
- gi No. 4226151
- Description: (AF125462) similar GAR1 related proteins (small nucleolar protein required for pre-rRNA splicing) [Caenorhabditis elegans]
- % Identity: 83.3
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 3315: from 21 to 32

Maximum Length Sequence:
related to:
Clone IDs:
240967
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3316
- Ceres seq_id 1576433
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3317
- Ceres seq_id 1576434
- Location of start within SEQ ID NO 3316: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3318
- Ceres seq_id 1576435
- Location of start within SEQ ID NO 3316: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3319
- Ceres seq_id 1576436
- Location of start within SEQ ID NO 3316: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SecE/Sec61-gamma subunits of protein translocation complex
- Location within SEQ ID NO 3319: from 10 to 66 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1504

- gi No. 585963
- Description: PROTEIN TRANSPORT PROTEIN SEC61 GAMMA SUBUNIT
- % Identity: 98.6
- Alignment Length: 69
- Location of Alignment in SEQ ID NO 3319: from 1 to 69

Maximum Length Sequence:
related to:
Clone IDs:
241045
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3320
- Ceres seq_id 1576437
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3321
- Ceres seq_id 1576438
- Location of start within SEQ ID NO 3320: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1505
- gi No. 1352427
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 1A (EIF-1A) (EIF-4C)
- % Identity: 96.2
- Alignment Length: 26
- Location of Alignment in SEQ ID NO 3321: from 33 to 58

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3322
- Ceres seq_id 1576439
- Location of start within SEQ ID NO 3320: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic initiation factor 1A
- Location within SEQ ID NO 3322: from 60 to 117 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1506
- gi No. 1352427
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 1A (EIF-1A) (EIF-4C)
- % Identity: 96.6
- Alignment Length: 58
- Location of Alignment in SEQ ID NO 3322: from 60 to 117

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3323
- Ceres seq_id 1576440
- Location of start within SEQ ID NO 3320: at 98 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1507
- gi No. 1352427
- Description: EUKARYOTIC TRANSLATION INITIATION FACTOR 1A (EIF-1A) (EIF-4C)
- % Identity: 96.2

- Alignment Length: 26
- Location of Alignment in SEQ ID NO 3323: from 1 to 26

Maximum Length Sequence:
related to:
Clone IDs:
241302
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3324
- Ceres seq_id 1576441
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3325
- Ceres seq_id 1576442
- Location of start within SEQ ID NO 3324: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KH domain
- Location within SEQ ID NO 3325: from 204 to 245 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3326
- Ceres seq_id 1576443
- Location of start within SEQ ID NO 3324: at 197 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KH domain
- Location within SEQ ID NO 3326: from 139 to 180 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
241341
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3327
- Ceres seq_id 1576444
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3328
- Ceres seq_id 1576445
- Location of start within SEQ ID NO 3327: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TCP-1/cpn60 chaperonin family
- Location within SEQ ID NO 3328: from 1 to 80 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3329
- Ceres seq_id 1576446
- Location of start within SEQ ID NO 3327: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TCP-1/cpn60 chaperonin family

- Location within SEQ ID NO 3329: from 1 to 66 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3330
- Ceres seq_id 1576447
- Location of start within SEQ ID NO 3327: at 478 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
242141
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3331
- Ceres seq_id 1576470
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3332
- Ceres seq_id 1576471
- Location of start within SEQ ID NO 3331: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 3332: from 1 to 55 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1508
- gi No. 2497903
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1752831|dbj|BAA14038.1| (D89931) metallothionein-like protein [Oryza sativa] >gi|1815628 (U43530) metallothionein-like type 2 [Oryza sativa]
- % Identity: 70.2
- Alignment Length: 47
- Location of Alignment in SEQ ID NO 3332: from 1 to 47

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3333
- Ceres seq_id 1576472
- Location of start within SEQ ID NO 3331: at 303 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 3333: from 7 to 141 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1509
- gi No. 4506701
- Description: ref|NP_001016.1|pRPS23| ribosomal protein S23 >gi|730647|sp|P39028|RS23_HUMAN 40S RIBOSOMAL PROTEIN S23 >gi|543449|pir||S41955 ribosomal protein S23 - rat >gi|631360|pir||S42105 ribosomal protein S23, cytosolic - human norvegicus]
- % Identity: 71.3
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 3333: from 1 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3334
- Ceres seq_id 1576473
- Location of start within SEQ ID NO 3331: at 321 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 3334: from 1 to 135 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1510
- gi No. 4506701
- Description: ref|NP_001016.1|pRPS23| ribosomal protein S23 >gi|730647|sp|P39028|RS23_HUMAN 40S RIBOSOMAL PROTEIN S23 >gi|543449|pir||S41955 ribosomal protein S23 - rat >gi|631360|pir||S42105 ribosomal protein S23, cytosolic - human norvegicus]
- % Identity: 71.3
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 3334: from 1 to 136

Maximum Length Sequence:
related to:
Clone IDs:
242383
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3335
- Ceres seq_id 1576485
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3336
- Ceres seq_id 1576486
- Location of start within SEQ ID NO 3335: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1511
- gi No. 112110
- Description: p23 protein - rat >gi|246868|bbs|90835 nucleic acid-binding protein p23=yeast ribosomal protein YL43 homolog [rats, liver, Peptide Partial, 28 aa]
- % Identity: 74.1
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3336: from 26 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3337
- Ceres seq_id 1576487
- Location of start within SEQ ID NO 3335: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1512
- gi No. 112110
- Description: p23 protein - rat >gi|246868|bbs|90835 nucleic acid-binding protein p23=yeast ribosomal protein YL43 homolog [rats, liver, Peptide Partial, 28 aa]
- % Identity: 74.1
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3337: from 2 to 28

Maximum Length Sequence:
related to:
Clone IDs:
242726
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3338
- Ceres seq_id 1576491
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3339
- Ceres seq_id 1576492
- Location of start within SEQ ID NO 3338: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3339: from 49 to 110 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3340
- Ceres seq_id 1576493
- Location of start within SEQ ID NO 3338: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
243106
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3341
- Ceres seq_id 1576494
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3342
- Ceres seq_id 1576495
- Location of start within SEQ ID NO 3341: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1513
- gi No. 2739044
- Description: (AF024651) polyphosphoinositide binding protein Sshlp [Glycine max]
- % Identity: 71
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3342: from 31 to 205

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3343
- Ceres seq_id 1576496
- Location of start within SEQ ID NO 3341: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1514

- gi No. 2739044
- Description: (AF024651) polyphosphoinositide binding protein Ssh1p [Glycine max]
- % Identity: 71
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3343: from 1 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3344
- Ceres seq_id 1576497
- Location of start within SEQ ID NO 3341: at 136 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1515
- gi No. 2739044
- Description: (AF024651) polyphosphoinositide binding protein Ssh1p [Glycine max]
- % Identity: 71
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3344: from 1 to 160

Maximum Length Sequence:
related to:
Clone IDs:
243377
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3345
- Ceres seq_id 1576508
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3346
- Ceres seq_id 1576509
- Location of start within SEQ ID NO 3345: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sperm histone P2
- Location within SEQ ID NO 3346: from 53 to 127 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3347
- Ceres seq_id 1576510
- Location of start within SEQ ID NO 3345: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3348
- Ceres seq_id 1576511
- Location of start within SEQ ID NO 3345: at 367 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:

related to:
Clone IDs:
243478
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3349
- Ceres seq_id 1576523
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3350
- Ceres seq_id 1576524
- Location of start within SEQ ID NO 3349: at 102 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 3350: from 34 to 82 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3351
- Ceres seq_id 1576525
- Location of start within SEQ ID NO 3349: at 123 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 3351: from 27 to 75 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3352
- Ceres seq_id 1576526
- Location of start within SEQ ID NO 3349: at 377 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
243744
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3353
- Ceres seq_id 1576535
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3354
- Ceres seq_id 1576536
- Location of start within SEQ ID NO 3353: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1516
- gi No. 1917019
- Description: (U92045) ribosomal protein S6 RPS6-1 [Zea mays]
- % Identity: 98.6
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 3354: from 56 to 203

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3355
- Ceres seq_id 1576537
- Location of start within SEQ ID NO 3353: at 204 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1517
- gi No. 1917019
- Description: (U92045) ribosomal protein S6 RPS6-1 [Zea mays]
- % Identity: 98.6
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 3355: from 18 to 165

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3356
- Ceres seq_id 1576538
- Location of start within SEQ ID NO 3353: at 231 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1518
- gi No. 1917019
- Description: (U92045) ribosomal protein S6 RPS6-1 [Zea mays]
- % Identity: 98.6
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 3356: from 9 to 156

Maximum Length Sequence:
related to:
Clone IDs:
244057

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3357
- Ceres seq_id 1576543

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3358
- Ceres seq_id 1576544
- Location of start within SEQ ID NO 3357: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 3358: from 1 to 58 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3359
- Ceres seq_id 1576545
- Location of start within SEQ ID NO 3357: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 3359: from 1 to 54 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3360
- Ceres seq_id 1576546
- Location of start within SEQ ID NO 3357: at 267 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
244364
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3361
- Ceres seq_id 1576564
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3362
- Ceres seq_id 1576565
- Location of start within SEQ ID NO 3361: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3363
- Ceres seq_id 1576566
- Location of start within SEQ ID NO 3361: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1519
- gi No. 3747044
- Description: (AF093537) blue copper protein [Zea mays]
- % Identity: 100
- Alignment Length: 108
- Location of Alignment in SEQ ID NO 3363: from 1 to 108

Maximum Length Sequence:
related to:
Clone IDs:
245016
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3364
- Ceres seq_id 1576575
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3365
- Ceres seq_id 1576576
- Location of start within SEQ ID NO 3364: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1520
- gi No. 1362192
- Description: ferredoxin--thioredoxin reductase (EC 1.18.-.-) variable chain - maize
- % Identity: 72.7
- Alignment Length: 22

- Location of Alignment in SEQ ID NO 3365: from 75 to 96

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3366
- Ceres seq_id 1576577
- Location of start within SEQ ID NO 3364: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3367
- Ceres seq_id 1576578
- Location of start within SEQ ID NO 3364: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1521
- gi No. 1362192
- Description: ferredoxin--thioredoxin reductase (EC 1.18.-.-) variable chain - maize
- % Identity: 72.7
- Alignment Length: 22
- Location of Alignment in SEQ ID NO 3367: from 66 to 87

Maximum Length Sequence:
related to:
Clone IDs:
245275
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3368
- Ceres seq_id 1576582
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3369
- Ceres seq_id 1576583
- Location of start within SEQ ID NO 3368: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1522
- gi No. 2191138
- Description: (AF007269) A_IG002N01.18 gene product [Arabidopsis thaliana]
- % Identity: 71.8
- Alignment Length: 110
- Location of Alignment in SEQ ID NO 3369: from 98 to 206

Maximum Length Sequence:
related to:
Clone IDs:
245771
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3370
- Ceres seq_id 1576592
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3371
- Ceres seq_id 1576593
- Location of start within SEQ ID NO 3370: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1523
- gi No. 2134209
- Description: protamine II-3 - painted turtle
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 3371: from 34 to 50

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3372
- Ceres seq_id 1576594
- Location of start within SEQ ID NO 3370: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3373
- Ceres seq_id 1576595
- Location of start within SEQ ID NO 3370: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
247175
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3374
- Ceres seq_id 1576618
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3375
- Ceres seq_id 1576619
- Location of start within SEQ ID NO 3374: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L34
- Location within SEQ ID NO 3375: from 129 to 163 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1524
- gi No. 132909
- Description: 50S RIBOSOMAL PROTEIN L34 >gi|71360|pir||R6PS34 ribosomal protein L34 - Pseudomonas putida >gi|45706|emb|CAA44414| (X62540) homologous to E.coli rpmH [Pseudomonas putida]
- % Identity: 76.7
- Alignment Length: 30
- Location of Alignment in SEQ ID NO 3375: from 134 to 163

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3376
- Ceres seq_id 1576620
- Location of start within SEQ ID NO 3374: at 27 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L34
- Location within SEQ ID NO 3376: from 121 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1525
- gi No. 132909
- Description: 50S RIBOSOMAL PROTEIN L34 >gi|71360|pir||R6PS34 ribosomal protein L34 - Pseudomonas putida >gi|45706|emb|CAA44414| (X62540) homologous to E.coli rpmH [Pseudomonas putida]
- % Identity: 76.7
- Alignment Length: 30
- Location of Alignment in SEQ ID NO 3376: from 126 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3377
- Ceres seq_id 1576621
- Location of start within SEQ ID NO 3374: at 51 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L34
- Location within SEQ ID NO 3377: from 113 to 147 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1526
- gi No. 132909
- Description: 50S RIBOSOMAL PROTEIN L34 >gi|71360|pir||R6PS34 ribosomal protein L34 - Pseudomonas putida >gi|45706|emb|CAA44414| (X62540) homologous to E.coli rpmH [Pseudomonas putida]
- % Identity: 76.7
- Alignment Length: 30
- Location of Alignment in SEQ ID NO 3377: from 118 to 147

Maximum Length Sequence:
related to:
Clone IDs:
257087
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3378
- Ceres seq_id 1576636
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3379
- Ceres seq_id 1576637
- Location of start within SEQ ID NO 3378: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1527
- gi No. 2493852
- Description: CYTOCHROME C OXIDASE POLYPEPTIDE VC >gi|1070356|emb|CAA92107| (Z68091) cytochrome c oxidase, Vc subunit [Hordeum vulgare]
- % Identity: 100
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3379: from 43 to 69

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3380

- Ceres seq_id 1576638
- Location of start within SEQ ID NO 3378: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3381
- Ceres seq_id 1576639
- Location of start within SEQ ID NO 3378: at 127 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1528
- gi No. 2493852
- Description: CYTOCHROME C OXIDASE POLYPEPTIDE VC >gi|1070356|emb|CAA92107| (Z68091) cytochrome c oxidase, Vc subunit [Hordeum vulgare]
- % Identity: 100
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3381: from 1 to 27

Maximum Length Sequence:
related to:
Clone IDs:
257278
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3382
- Ceres seq_id 1576648
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3383
- Ceres seq_id 1576649
- Location of start within SEQ ID NO 3382: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3384
- Ceres seq_id 1576650
- Location of start within SEQ ID NO 3382: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3385
- Ceres seq_id 1576651
- Location of start within SEQ ID NO 3382: at 230 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1529
- gi No. 3717989
- Description: (AJ005900) vag2 [Nicotiana tabacum]
- % Identity: 72.6

- Alignment Length: 84
- Location of Alignment in SEQ ID NO 3385: from 3 to 86

Maximum Length Sequence:
related to:
Clone IDs:
257780
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3386
- Ceres seq_id 1576665
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3387
- Ceres seq_id 1576666
- Location of start within SEQ ID NO 3386: at 351 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S17
- Location within SEQ ID NO 3387: from 1 to 57 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1530
- gi No. 133867
- Description: 40S RIBOSOMAL PROTEIN S11 >gi|82722|pir||S16577 ribosomal protein S11 - maize >gi|22470|emb|CAA39438| (X55967) ribosomal protein S11 [Zea mays]
- % Identity: 92.5
- Alignment Length: 133
- Location of Alignment in SEQ ID NO 3387: from 1 to 57

Maximum Length Sequence:
related to:
Clone IDs:
257928
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3388
- Ceres seq_id 1576671
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3389
- Ceres seq_id 1576672
- Location of start within SEQ ID NO 3388: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3390
- Ceres seq_id 1576673
- Location of start within SEQ ID NO 3388: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3391
- Ceres seq_id 1576674
- Location of start within SEQ ID NO 3388: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1531
- gi No. 2258365
- Description: (AF007261) NADH dehydrogenase, subunit 10 [Reclinomonas americana]
- % Identity: 76.6
- Alignment Length: 47
- Location of Alignment in SEQ ID NO 3391: from 46 to 92

Maximum Length Sequence:
related to:
Clone IDs:
258447
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3392
- Ceres seq_id 1576682

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3393
- Ceres seq_id 1576683
- Location of start within SEQ ID NO 3392: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1532
- gi No. 4914414
- Description: (AL050352) Ca2+-transporting ATPase-like protein [Arabidopsis thaliana]
- % Identity: 70
- Alignment Length: 50
- Location of Alignment in SEQ ID NO 3393: from 18 to 65

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3394
- Ceres seq_id 1576684
- Location of start within SEQ ID NO 3392: at 457 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S17
- Location within SEQ ID NO 3394: from 1 to 63 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1533
- gi No. 133867
- Description: 40S RIBOSOMAL PROTEIN S11 >gi|82722|pir||S16577 ribosomal protein S11 - maize >gi|22470|emb|CAA39438| (X55967) ribosomal protein S11 [Zea mays]
- % Identity: 99.2
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3394: from 1 to 79

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3395
- Ceres seq_id 1576685
- Location of start within SEQ ID NO 3392: at 558 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
260368
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3396
- Ceres seq_id 1576713
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3397
- Ceres seq_id 1576714
- Location of start within SEQ ID NO 3396: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Zinc finger, C2H2 type
- Location within SEQ ID NO 3397: from 25 to 47 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1534
- gi No. 1361986
- Description: finger protein 2 - Arabidopsis thaliana >gi|790675 (L39645) zinc finger protein [Arabidopsis thaliana]
- % Identity: 70.3
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 3397: from 15 to 51

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3398
- Ceres seq_id 1576715
- Location of start within SEQ ID NO 3396: at 347 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3399
- Ceres seq_id 1576716
- Location of start within SEQ ID NO 3396: at 460 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
261335
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3400
- Ceres seq_id 1576730
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3401
- Ceres seq_id 1576731
- Location of start within SEQ ID NO 3400: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein

- Location within SEQ ID NO 3401: from 28 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1535
- gi No. 4755087
- Description: (AF031083) aluminum-induced protein; Al-induced protein [Zea mays]
- % Identity: 92.2
- Alignment Length: 115
- Location of Alignment in SEQ ID NO 3401: from 1 to 115

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3402
- Ceres seq_id 1576732
- Location of start within SEQ ID NO 3400: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein
- Location within SEQ ID NO 3402: from 73 to 140 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1536
- gi No. 4755087
- Description: (AF031083) aluminum-induced protein; Al-induced protein [Zea mays]
- % Identity: 92.4
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 3402: from 71 to 147

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3403
- Ceres seq_id 1576733
- Location of start within SEQ ID NO 3400: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 11-S plant seed storage protein
- Location within SEQ ID NO 3403: from 55 to 122 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1537
- gi No. 4755087
- Description: (AF031083) aluminum-induced protein; Al-induced protein [Zea mays]
- % Identity: 92.4
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 3403: from 53 to 129

Maximum Length Sequence:
related to:
Clone IDs:
261718
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3404
- Ceres seq_id 1576754
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3405
- Ceres seq_id 1576755
- Location of start within SEQ ID NO 3404: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1538
- gi No. 1710509
- Description: 60S RIBOSOMAL PROTEIN L18A
- % Identity: 87.8
- Alignment Length: 90
- Location of Alignment in SEQ ID NO 3405: from 69 to 158

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3406
- Ceres seq_id 1576756
- Location of start within SEQ ID NO 3404: at 175 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1539
- gi No. 1710509
- Description: 60S RIBOSOMAL PROTEIN L18A
- % Identity: 87.8
- Alignment Length: 90
- Location of Alignment in SEQ ID NO 3406: from 41 to 130

Maximum Length Sequence:
related to:
Clone IDs:
261748

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3407
- Ceres seq_id 1576757

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3408
- Ceres seq_id 1576758
- Location of start within SEQ ID NO 3407: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Sm protein
- Location within SEQ ID NO 3408: from 94 to 155 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3409
- Ceres seq_id 1576759
- Location of start within SEQ ID NO 3407: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Sm protein
- Location within SEQ ID NO 3409: from 35 to 96 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
261831

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3410
- Ceres seq_id 1576763

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3411
- Ceres seq_id 1576764
- Location of start within SEQ ID NO 3410: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3412
- Ceres seq_id 1576765
- Location of start within SEQ ID NO 3410: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1540
- gi No. 132933
- Description: 60S RIBOSOMAL PROTEIN L37 (YL35)
- % Identity: 77.8
- Alignment Length: 27
- Location of Alignment in SEQ ID NO 3412: from 26 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3413
- Ceres seq_id 1576766
- Location of start within SEQ ID NO 3410: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
262995
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3414
- Ceres seq_id 1576795

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3415
- Ceres seq_id 1576796
- Location of start within SEQ ID NO 3414: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1541
- gi No. 2459425
- Description: (AC002332) plastid protein [Arabidopsis thaliana]
- % Identity: 73.8
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3415: from 33 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3416
- Ceres seq_id 1576797
- Location of start within SEQ ID NO 3414: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3417
- Ceres seq_id 1576798
- Location of start within SEQ ID NO 3414: at 49 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1542
- gi No. 2459425
- Description: (AC002332) plastid protein [Arabidopsis thaliana]
- % Identity: 73.8
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3417: from 17 to 139

Maximum Length Sequence:
related to:
Clone IDs:
263037
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3418
- Ceres seq_id 1576799
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3419
- Ceres seq_id 1576800
- Location of start within SEQ ID NO 3418: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3420
- Ceres seq_id 1576801
- Location of start within SEQ ID NO 3418: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA polymerases N / 8 kDa subunit
- Location within SEQ ID NO 3420: from 44 to 103 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1543
- gi No. 2833375
- Description: DNA-DIRECTED RNA POLYMERASE II 8.2 KD POLYPEPTIDE (RPB10) (RP10) (ABC10) >gi|533690 (U12133) RNA polymerase II subunit RPB10 homolog; similar to yeast RNA polymerase II subunit RPB10, Swiss-Prot Accession Number P22139 [Brassica napus]
- % Identity: 88.7
- Alignment Length: 71
- Location of Alignment in SEQ ID NO 3420: from 44 to 114

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3421
- Ceres seq_id 1576802
- Location of start within SEQ ID NO 3418: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA polymerases N / 8 kDa subunit
- Location within SEQ ID NO 3421: from 1 to 60 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1544
- gi No. 2833375
- Description: DNA-DIRECTED RNA POLYMERASE II 8.2 KD POLYPEPTIDE (RPB10) (RP10) (ABC10) >gi|533690 (U12133) RNA polymerase II subunit RPB10 homolog; similar to yeast RNA polymerase II subunit RPB10, Swiss-Prot Accession Number P22139 [Brassica napus]
- % Identity: 88.7
- Alignment Length: 71
- Location of Alignment in SEQ ID NO 3421: from 1 to 71

Maximum Length Sequence:
related to:
Clone IDs:
263051
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3422
- Ceres seq_id 1576803
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3423
- Ceres seq_id 1576804
- Location of start within SEQ ID NO 3422: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3424
- Ceres seq_id 1576805
- Location of start within SEQ ID NO 3422: at 83 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 3424: from 6 to 173 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1545
- gi No. 1084454
- Description: peptidylprolyl isomerase (EC 5.2.1.8) Cyp1 - rice >gi|600765 (L29471) cyclophilin 1 [Oryza sativa]
- % Identity: 87.4
- Alignment Length: 174
- Location of Alignment in SEQ ID NO 3424: from 2 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3425
- Ceres seq_id 1576806
- Location of start within SEQ ID NO 3422: at 149 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 3425: from 1 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1546
- gi No. 1084454
- Description: peptidylprolyl isomerase (EC 5.2.1.8) Cyp1 - rice >gi|600765 (L29471) cyclophilin 1 [Oryza sativa]
- % Identity: 87.4
- Alignment Length: 174
- Location of Alignment in SEQ ID NO 3425: from 1 to 153

Maximum Length Sequence:
related to:
Clone IDs:
263375
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3426
- Ceres seq_id 1576807
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3427
- Ceres seq_id 1576808
- Location of start within SEQ ID NO 3426: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1547
- gi No. 4098250
- Description: (U76611) similar to Solanum tuberosum ci21A gene product encoded by the sequence presented in GenBank Accession Number U76610 [Solanum tuberosum]
- % Identity: 75.8
- Alignment Length: 99
- Location of Alignment in SEQ ID NO 3427: from 1 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3428
- Ceres seq_id 1576809
- Location of start within SEQ ID NO 3426: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1548
- gi No. 4098250
- Description: (U76611) similar to Solanum tuberosum ci21A gene product encoded by the sequence presented in GenBank Accession Number U76610 [Solanum tuberosum]
- % Identity: 75.8
- Alignment Length: 99
- Location of Alignment in SEQ ID NO 3428: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
263742
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3429
- Ceres seq_id 1576816
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3430
- Ceres seq_id 1576817

- Location of start within SEQ ID NO 3429: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3431
- Ceres seq_id 1576818
- Location of start within SEQ ID NO 3429: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 3431: from 34 to 100 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3432
- Ceres seq_id 1576819
- Location of start within SEQ ID NO 3429: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
264309
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3433
- Ceres seq_id 1576842
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3434
- Ceres seq_id 1576843
- Location of start within SEQ ID NO 3433: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3435
- Ceres seq_id 1576844
- Location of start within SEQ ID NO 3433: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1549
- gi No. 4539386
- Description: (AL035526) extensin-like protein [Arabidopsis thaliana]
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3435: from 128 to 141

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3436

- Ceres seq_id 1576845
- Location of start within SEQ ID NO 3433: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1550
- gi No. 4539386
- Description: (AL035526) extensin-like protein [Arabidopsis thaliana]
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3436: from 92 to 105

Maximum Length Sequence:
related to:
Clone IDs:
264720
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3437
- Ceres seq_id 1576855
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3438
- Ceres seq_id 1576856
- Location of start within SEQ ID NO 3437: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3439
- Ceres seq_id 1576857
- Location of start within SEQ ID NO 3437: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Voltage gated chloride channels
- Location within SEQ ID NO 3439: from 46 to 129 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3440
- Ceres seq_id 1576858
- Location of start within SEQ ID NO 3437: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Voltage gated chloride channels
- Location within SEQ ID NO 3440: from 24 to 107 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
265197
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3441
- Ceres seq_id 1576876

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3442
- Ceres seq_id 1576877
- Location of start within SEQ ID NO 3441: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3442: from 151 to 245 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1551
- gi No. 1658315
- Description: (Y08988) osr40g3 [Oryza sativa]
- % Identity: 77.2
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 3442: from 5 to 129

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3443
- Ceres seq_id 1576878
- Location of start within SEQ ID NO 3441: at 205 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3443: from 116 to 210 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1552
- gi No. 1658315
- Description: (Y08988) osr40g3 [Oryza sativa]
- % Identity: 77.2
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 3443: from 1 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3444
- Ceres seq_id 1576879
- Location of start within SEQ ID NO 3441: at 301 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3444: from 84 to 178 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1553
- gi No. 1658315
- Description: (Y08988) osr40g3 [Oryza sativa]
- % Identity: 77.2
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 3444: from 1 to 62

Maximum Length Sequence:
related to:
Clone IDs:
265543
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3445
- Ceres seq_id 1576888

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3446
- Ceres seq_id 1576889
- Location of start within SEQ ID NO 3445: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3447
- Ceres seq_id 1576890
- Location of start within SEQ ID NO 3445: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Translationally controlled tumor protein
- Location within SEQ ID NO 3447: from 1 to 133 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1554
- gi No. 549063
- Description: TRANSLATIONALLY CONTROLLED TUMOR PROTEIN HOMOLOG (TCTP) >gi|1072464|pir||A38958 IgE-dependent histamine-releasing factor homolog - rice >gi|303835|dbj|BAA02151| (D12626) 21kd polypeptide [Oryza sativa]
- % Identity: 73.9
- Alignment Length: 139
- Location of Alignment in SEQ ID NO 3447: from 1 to 133

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3448
- Ceres seq_id 1576891
- Location of start within SEQ ID NO 3445: at 488 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
266028
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3449
- Ceres seq_id 1576896

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3450
- Ceres seq_id 1576897
- Location of start within SEQ ID NO 3449: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3451
- Ceres seq_id 1576898
- Location of start within SEQ ID NO 3449: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 3451: from 20 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1555
- gi No. 4995921
- Description: (AJ006708) HMG1 protein [Zea mays]
- % Identity: 99.2
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 3451: from 1 to 123

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3452
- Ceres seq_id 1576899
- Location of start within SEQ ID NO 3449: at 320 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1556
- gi No. 4995921
- Description: (AJ006708) HMG1 protein [Zea mays]
- % Identity: 99.2
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 3452: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
272573
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3453
- Ceres seq_id 1576923
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3454
- Ceres seq_id 1576924
- Location of start within SEQ ID NO 3453: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3455
- Ceres seq_id 1576925
- Location of start within SEQ ID NO 3453: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1557
- gi No. 1084968
- Description: regulatory protein - Emericella nidulans >gi|540284 (L36341) regulatory protein [Emericella nidulans]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3455: from 107 to 117

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3456
- Ceres seq_id 1576926
- Location of start within SEQ ID NO 3453: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1558
- gi No. 1084968
- Description: regulatory protein - Emericella nidulans >gi|540284 (L36341) regulatory protein [Emericella nidulans]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3456: from 70 to 80

Maximum Length Sequence:
related to:
Clone IDs:
272605

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3457
- Ceres seq_id 1576935

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3458
- Ceres seq_id 1576936
- Location of start within SEQ ID NO 3457: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3459
- Ceres seq_id 1576937
- Location of start within SEQ ID NO 3457: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1559
- gi No. 2493053
- Description: ATP SYNTHASE EPSILON CHAIN, MITOCHONDRIAL >gi|639793 (L39120) mitochondrial F1F0 ATP synthase epsilon subunit [Zea mays]
- % Identity: 97.1
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 3459: from 34 to 103

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3460
- Ceres seq_id 1576938
- Location of start within SEQ ID NO 3457: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1560
- gi No. 2493053
- Description: ATP SYNTHASE EPSILON CHAIN, MITOCHONDRIAL >gi|639793 (L39120) mitochondrial F1F0 ATP synthase epsilon subunit [Zea mays]
- % Identity: 97.1
- Alignment Length: 70

- Location of Alignment in SEQ ID NO 3460: from 1 to 70

Maximum Length Sequence:
related to:
Clone IDs:
272643
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3461
- Ceres seq_id 1576943
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3462
- Ceres seq_id 1576944
- Location of start within SEQ ID NO 3461: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1561
- gi No. 439289
- Description: (Z26645) verprolin [Saccharomyces cerevisiae]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3462: from 84 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3463
- Ceres seq_id 1576945
- Location of start within SEQ ID NO 3461: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1562
- gi No. 4507909
- Description: ref|NP_000368.1|pWAS| Wiskott-Aldrich syndrome (ecezema-thrombocytopenia) >gi|1722836|sp|P42768|WASP_HUMAN WISKOTT-ALDRICH SYNDROME PROTEIN (WASP) >gi|2136367|pir||I38931 Wiskott-Aldrich syndrome protein WASP - human >gi|695151 (U12707)
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 3463: from 135 to 151

Maximum Length Sequence:
related to:
Clone IDs:
302917
272664
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3464
- Ceres seq_id 1576950
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3465
- Ceres seq_id 1576951
- Location of start within SEQ ID NO 3464: at 144 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3465: from 51 to 223 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3466
- Ceres seq_id 1576952
- Location of start within SEQ ID NO 3464: at 375 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3466: from 1 to 146 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
272681
324869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3467
- Ceres seq_id 1576959
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3468
- Ceres seq_id 1576960
- Location of start within SEQ ID NO 3467: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 3468: from 57 to 135 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1563
- gi No. 5262856
- Description: (AJ238095) Lsm3 protein [Homo sapiens]
- % Identity: 83.9
- Alignment Length: 87
- Location of Alignment in SEQ ID NO 3468: from 52 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3469
- Ceres seq_id 1576961
- Location of start within SEQ ID NO 3467: at 128 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 3469: from 15 to 93 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1564
- gi No. 5262856
- Description: (AJ238095) Lsm3 protein [Homo sapiens]
- % Identity: 83.9
- Alignment Length: 87
- Location of Alignment in SEQ ID NO 3469: from 10 to 96

Maximum Length Sequence:
related to:
Clone IDs:
272735

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3470
- Ceres seq_id 1576970

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3471
- Ceres seq_id 1576971
- Location of start within SEQ ID NO 3470: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1565
- gi No. 2564066
- Description: (D45900) LEDI-3 protein [Lithospermum erythrorhizon]
- % Identity: 72.1
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 3471: from 44 to 111

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3472
- Ceres seq_id 1576972
- Location of start within SEQ ID NO 3470: at 380 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1566
- gi No. 2564066
- Description: (D45900) LEDI-3 protein [Lithospermum erythrorhizon]
- % Identity: 76.5
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 3472: from 1 to 64

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3473
- Ceres seq_id 1576973
- Location of start within SEQ ID NO 3470: at 383 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1567
- gi No. 2564066
- Description: (D45900) LEDI-3 protein [Lithospermum erythrorhizon]
- % Identity: 76.5
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 3473: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
273017

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3474
- Ceres seq_id 1576996

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3475
- Ceres seq_id 1576997
- Location of start within SEQ ID NO 3474: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 3475: from 28 to 150 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
273031
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3476
- Ceres seq_id 1576998
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3477
- Ceres seq_id 1576999
- Location of start within SEQ ID NO 3476: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3478
- Ceres seq_id 1577000
- Location of start within SEQ ID NO 3476: at 360 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3478: from 1 to 83 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1568
- gi No. 5381319
- Description: (AF091621) ubiquitin-conjugating enzyme E2 [Catharanthus roseus]
- % Identity: 94.6
- Alignment Length: 112
- Location of Alignment in SEQ ID NO 3478: from 1 to 88

Maximum Length Sequence:
related to:
Clone IDs:
273175
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3479
- Ceres seq_id 1577001
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3480
- Ceres seq_id 1577002
- Location of start within SEQ ID NO 3479: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3481
- Ceres seq_id 1577003

- Location of start within SEQ ID NO 3479: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1569
- gi No. 2134209
- Description: protamine II-3 - painted turtle
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3481: from 49 to 66

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3482
- Ceres seq_id 1577004
- Location of start within SEQ ID NO 3479: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
273508

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3483
- Ceres seq_id 1577048

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3484
- Ceres seq_id 1577049
- Location of start within SEQ ID NO 3483: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 3484: from 24 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1570
- gi No. 2431771
- Description: (U62753) acidic ribosomal protein P2b [Zea mays]
- % Identity: 93.8
- Alignment Length: 96
- Location of Alignment in SEQ ID NO 3484: from 24 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3485
- Ceres seq_id 1577050
- Location of start within SEQ ID NO 3483: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 3485: from 1 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1571
- gi No. 2431771
- Description: (U62753) acidic ribosomal protein P2b [Zea mays]
- % Identity: 93.8

- Alignment Length: 96
- Location of Alignment in SEQ ID NO 3485: from 1 to 96

Maximum Length Sequence:
related to:
Clone IDs:
273567
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3486
- Ceres seq_id 1577058
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3487
- Ceres seq_id 1577059
- Location of start within SEQ ID NO 3486: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3487: from 91 to 145 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1572
- gi No. 3077806
- Description: (AB013076) loricrin [Homo sapiens]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3487: from 197 to 207

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3488
- Ceres seq_id 1577060
- Location of start within SEQ ID NO 3486: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3488: from 60 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1573
- gi No. 3077806
- Description: (AB013076) loricrin [Homo sapiens]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3488: from 166 to 176

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3489
- Ceres seq_id 1577061
- Location of start within SEQ ID NO 3486: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3489: from 56 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1574
- gi No. 3077806
- Description: (AB013076) loricrin [Homo sapiens]
- % Identity: 81.8

- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3489: from 162 to 172

Maximum Length Sequence:
related to:
Clone IDs:
273577
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3490
- Ceres seq_id 1577062
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3491
- Ceres seq_id 1577063
- Location of start within SEQ ID NO 3490: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3492
- Ceres seq_id 1577064
- Location of start within SEQ ID NO 3490: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1575
- gi No. 3264600
- Description: (AF057522) hypoxically induced transcript 2 [Zea mays]
- % Identity: 72.6
- Alignment Length: 73
- Location of Alignment in SEQ ID NO 3492: from 143 to 215

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3493
- Ceres seq_id 1577065
- Location of start within SEQ ID NO 3490: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1576
- gi No. 3264600
- Description: (AF057522) hypoxically induced transcript 2 [Zea mays]
- % Identity: 72.6
- Alignment Length: 73
- Location of Alignment in SEQ ID NO 3493: from 92 to 164

Maximum Length Sequence:
related to:
Clone IDs:
273675
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3494
- Ceres seq_id 1577073
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3495
- Ceres seq_id 1577074

- Location of start within SEQ ID NO 3494: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3496
- Ceres seq_id 1577075
- Location of start within SEQ ID NO 3494: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1577
- gi No. 1173350
- Description: DNA BINDING PROTEIN S1FA
- % Identity: 83.8
- Alignment Length: 74
- Location of Alignment in SEQ ID NO 3496: from 7 to 80

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3497
- Ceres seq_id 1577076
- Location of start within SEQ ID NO 3494: at 447 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
273782
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3498
- Ceres seq_id 1577089
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3499
- Ceres seq_id 1577090
- Location of start within SEQ ID NO 3498: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1578
- gi No. 1173137
- Description: DNA-DIRECTED RNA POLYMERASE II 19 KD POLYPEPTIDE (RNA POLYMERASE II SUBUNIT 5) >gi|322700|pir||B44457 RNA polymerase II fifth largest subunit - Glycine max=soybeans >gi|170052 (M90504) RNA polymerase II [Glycine max]
- % Identity: 70.1
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3499: from 35 to 164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3500
- Ceres seq_id 1577091
- Location of start within SEQ ID NO 3498: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1579
- gi No. 1173137
- Description: DNA-DIRECTED RNA POLYMERASE II 19 KD POLYPEPTIDE (RNA POLYMERASE II SUBUNIT 5) >gi|322700|pir||B44457 RNA polymerase II fifth largest subunit - Glycine max=soybeans >gi|170052 (M90504) RNA polymerase II [Glycine max]
- % Identity: 70.1
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3500: from 1 to 130

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3501
- Ceres seq_id 1577092
- Location of start within SEQ ID NO 3498: at 134 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1580
- gi No. 1173137
- Description: DNA-DIRECTED RNA POLYMERASE II 19 KD POLYPEPTIDE (RNA POLYMERASE II SUBUNIT 5) >gi|322700|pir||B44457 RNA polymerase II fifth largest subunit - Glycine max=soybeans >gi|170052 (M90504) RNA polymerase II [Glycine max]
- % Identity: 70.1
- Alignment Length: 176
- Location of Alignment in SEQ ID NO 3501: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
274081

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3502
- Ceres seq_id 1577122

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3503
- Ceres seq_id 1577123
- Location of start within SEQ ID NO 3502: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- KOW motif
- Location within SEQ ID NO 3503: from 78 to 133 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3504
- Ceres seq_id 1577124
- Location of start within SEQ ID NO 3502: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- KOW motif
- Location within SEQ ID NO 3504: from 75 to 130 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3505
- Ceres seq_id 1577125
- Location of start within SEQ ID NO 3502: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KOW motif
- Location within SEQ ID NO 3505: from 68 to 123 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274123
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3506
- Ceres seq_id 1577130
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3507
- Ceres seq_id 1577131
- Location of start within SEQ ID NO 3506: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3508
- Ceres seq_id 1577132
- Location of start within SEQ ID NO 3506: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1581
- gi No. 2244825
- Description: (Z97336) light induced protein like [Arabidopsis thaliana]
- % Identity: 81.4
- Alignment Length: 97
- Location of Alignment in SEQ ID NO 3508: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
274136
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3509
- Ceres seq_id 1577133
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3510
- Ceres seq_id 1577134
- Location of start within SEQ ID NO 3509: at 199 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S17
- Location within SEQ ID NO 3510: from 6 to 60 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3511
- Ceres seq_id 1577135
- Location of start within SEQ ID NO 3509: at 235 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3512
- Ceres seq_id 1577136
- Location of start within SEQ ID NO 3509: at 517 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274201
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3513
- Ceres seq_id 1577137
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3514
- Ceres seq_id 1577138
- Location of start within SEQ ID NO 3513: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1582
- gi No. 5006857
- Description: (AF145730) homeodomain leucine zipper protein [Oryza sativa]
- % Identity: 75
- Alignment Length: 37
- Location of Alignment in SEQ ID NO 3514: from 1 to 30

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3515
- Ceres seq_id 1577139
- Location of start within SEQ ID NO 3513: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1583
- gi No. 5006857
- Description: (AF145730) homeodomain leucine zipper protein [Oryza sativa]
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3515: from 88 to 105

Maximum Length Sequence:

related to:
Clone IDs:
274360
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3516
- Ceres seq_id 1577146
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3517
- Ceres seq_id 1577147
- Location of start within SEQ ID NO 3516: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 3
- Location within SEQ ID NO 3517: from 68 to 134 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3518
- Ceres seq_id 1577148
- Location of start within SEQ ID NO 3516: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolase family 3
- Location within SEQ ID NO 3518: from 49 to 115 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274426
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3519
- Ceres seq_id 1577163
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3520
- Ceres seq_id 1577164
- Location of start within SEQ ID NO 3519: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.
- Location within SEQ ID NO 3520: from 167 to 228 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3521
- Ceres seq_id 1577165
- Location of start within SEQ ID NO 3519: at 393 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.
- Location within SEQ ID NO 3521: from 37 to 98 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3522
- Ceres seq_id 1577166
- Location of start within SEQ ID NO 3519: at 420 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.
- Location within SEQ ID NO 3522: from 28 to 89 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274520
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3523
- Ceres seq_id 1577167
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3524
- Ceres seq_id 1577168
- Location of start within SEQ ID NO 3523: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1584
- gi No. 585551
- Description: NUCLEOSIDE DIPHOSPHATE KINASE I (NDK I) (NDP KINASE I) >gi|629798|pir||S43330 nucleoside-diphosphate kinase (EC 2.7.4.6) - rice >gi|303849|dbj|BAA03798| (D16292) nucleoside diphosphate kinase [Oryza sativa]
- % Identity: 86
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 3524: from 47 to 96

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3525
- Ceres seq_id 1577169
- Location of start within SEQ ID NO 3523: at 358 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Nucleoside diphosphate kinases
- Location within SEQ ID NO 3525: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1585
- gi No. 585551
- Description: NUCLEOSIDE DIPHOSPHATE KINASE I (NDK I) (NDP KINASE I) >gi|629798|pir||S43330 nucleoside-diphosphate kinase (EC 2.7.4.6) - rice >gi|303849|dbj|BAA03798| (D16292) nucleoside diphosphate kinase [Oryza sativa]
- % Identity: 86
- Alignment Length: 150
- Location of Alignment in SEQ ID NO 3525: from 1 to 77

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3526
- Ceres seq_id 1577170
- Location of start within SEQ ID NO 3523: at 476 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274576
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3527
- Ceres seq_id 1577175
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3528
- Ceres seq_id 1577176
- Location of start within SEQ ID NO 3527: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3529
- Ceres seq_id 1577177
- Location of start within SEQ ID NO 3527: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1586
- gi No. 1296429
- Description: (L77967) small proline-rich protein with paired repeat [Ovis aries]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3529: from 22 to 32

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3530
- Ceres seq_id 1577178
- Location of start within SEQ ID NO 3527: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274696
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3531
- Ceres seq_id 1577198
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3532
- Ceres seq_id 1577199
- Location of start within SEQ ID NO 3531: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3533
- Ceres seq_id 1577200
- Location of start within SEQ ID NO 3531: at 151 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Skp1 family
- Location within SEQ ID NO 3533: from 1 to 60 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3534
- Ceres seq_id 1577201
- Location of start within SEQ ID NO 3531: at 351 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
274739
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3535
- Ceres seq_id 1577206
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3536
- Ceres seq_id 1577207
- Location of start within SEQ ID NO 3535: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1587
- gi No. 1172818
- Description: 40S RIBOSOMAL PROTEIN S16 >gi|538428 (L36313) ribosomal protein S16 [Oryza sativa] >gi|1096552|prf||2111468A ribosomal protein S16 [Oryza sativa]
- % Identity: 89.7
- Alignment Length: 58
- Location of Alignment in SEQ ID NO 3536: from 68 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3537
- Ceres seq_id 1577209
- Location of start within SEQ ID NO 3535: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1588
- gi No. 1172818

- Description: 40S RIBOSOMAL PROTEIN S16 >gi|538428 (L36313) ribosomal protein S16 [Oryza sativa] >gi|1096552|prf||2111468A ribosomal protein S16 [Oryza sativa]
- % Identity: 89.7
- Alignment Length: 58
- Location of Alignment in SEQ ID NO 3537: from 29 to 86

Maximum Length Sequence:
related to:
Clone IDs:
275391
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3538
- Ceres seq_id 1577252
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3539
- Ceres seq_id 1577253
- Location of start within SEQ ID NO 3538: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1589
- gi No. 2665536
- Description: (AF027808) HCF106 precursor protein [Zea mays]
- % Identity: 90.5
- Alignment Length: 179
- Location of Alignment in SEQ ID NO 3539: from 33 to 209

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3540
- Ceres seq_id 1577254
- Location of start within SEQ ID NO 3538: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1590
- gi No. 2665536
- Description: (AF027808) HCF106 precursor protein [Zea mays]
- % Identity: 90.5
- Alignment Length: 179
- Location of Alignment in SEQ ID NO 3540: from 1 to 134

Maximum Length Sequence:
related to:
Clone IDs:
275597
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3541
- Ceres seq_id 1577263
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3542
- Ceres seq_id 1577264
- Location of start within SEQ ID NO 3541: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3543
- Ceres seq_id 1577265
- Location of start within SEQ ID NO 3541: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynein light chain type 1
- Location within SEQ ID NO 3543: from 84 to 135 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
276166
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3544
- Ceres seq_id 1577296
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3545
- Ceres seq_id 1577297
- Location of start within SEQ ID NO 3544: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1591
- gi No. 1181331
- Description: (X77569) calnexin [Zea mays]
- % Identity: 94.7
- Alignment Length: 38
- Location of Alignment in SEQ ID NO 3545: from 1 to 38

Maximum Length Sequence:
related to:
Clone IDs:
276284
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3546
- Ceres seq_id 1577302
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3547
- Ceres seq_id 1577303
- Location of start within SEQ ID NO 3546: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3548
- Ceres seq_id 1577304
- Location of start within SEQ ID NO 3546: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3549
- Ceres seq_id 1577305

- Location of start within SEQ ID NO 3546: at 124 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3549: from 20 to 124 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1592
- gi No. 100214
- Description: extensin class I (clone wY) - tomato >gi|1345535|emb|CAA39213| (X55682) extensin (class I) [Lycopersicon esculentum]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 3549: from 23 to 34

Maximum Length Sequence:
related to:
Clone IDs:
276353
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3550
- Ceres seq_id 1577310
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3551
- Ceres seq_id 1577311
- Location of start within SEQ ID NO 3550: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3552
- Ceres seq_id 1577312
- Location of start within SEQ ID NO 3550: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3553
- Ceres seq_id 1577313
- Location of start within SEQ ID NO 3550: at 172 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATP synthase subunit C
- Location within SEQ ID NO 3553: from 23 to 85 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1593
- gi No. 4874301
- Description: (AC006053) proton-ATPase-like protein [Arabidopsis thaliana]
- % Identity: 89.3
- Alignment Length: 121
- Location of Alignment in SEQ ID NO 3553: from 5 to 125

Maximum Length Sequence:
related to:
Clone IDs:
276371
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3554
- Ceres seq_id 1577318
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3555
- Ceres seq_id 1577319
- Location of start within SEQ ID NO 3554: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3556
- Ceres seq_id 1577320
- Location of start within SEQ ID NO 3554: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3557
- Ceres seq_id 1577321
- Location of start within SEQ ID NO 3554: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 3557: from 18 to 128 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1594
- gi No. 100883
- Description: heat shock protein 17.2 - maize
>gi|22335|emb|CAA46641| (X65725) heat shock protein 17.2 [Zea mays]
- % Identity: 85.5
- Alignment Length: 145
- Location of Alignment in SEQ ID NO 3557: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
276495
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3558
- Ceres seq_id 1577346
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3559
- Ceres seq_id 1577347
- Location of start within SEQ ID NO 3558: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3559: from 22 to 77 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3560
- Ceres seq_id 1577348
- Location of start within SEQ ID NO 3558: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3560: from 13 to 68 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3561
- Ceres seq_id 1577349
- Location of start within SEQ ID NO 3558: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
276634
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3562
- Ceres seq_id 1577353
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3563
- Ceres seq_id 1577354
- Location of start within SEQ ID NO 3562: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3564
- Ceres seq_id 1577355
- Location of start within SEQ ID NO 3562: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3565
- Ceres seq_id 1577356
- Location of start within SEQ ID NO 3562: at 296 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1595
- gi No. 1103318
- Description: (X78818) casein kinase I [Arabidopsis thaliana] >gi|2244791|emb|CAB10213.1| (Z97336) casein kinase I [Arabidopsis thaliana]
- % Identity: 80

- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3565: from 1 to 20

Maximum Length Sequence:
related to:
Clone IDs:
276903
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3566
- Ceres seq_id 1577374
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3567
- Ceres seq_id 1577375
- Location of start within SEQ ID NO 3566: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 3567: from 13 to 144 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3568
- Ceres seq_id 1577376
- Location of start within SEQ ID NO 3566: at 168 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 3568: from 1 to 89 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3569
- Ceres seq_id 1577377
- Location of start within SEQ ID NO 3566: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- GMC oxidoreductases
- Location within SEQ ID NO 3569: from 1 to 83 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
276985
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3570
- Ceres seq_id 1577378
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3571
- Ceres seq_id 1577379
- Location of start within SEQ ID NO 3570: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peroxidase

- Location within SEQ ID NO 3571: from 1 to 148 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3572
- Ceres seq_id 1577380
- Location of start within SEQ ID NO 3570: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3572: from 23 to 117 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3573
- Ceres seq_id 1577381
- Location of start within SEQ ID NO 3570: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
277001
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3574
- Ceres seq_id 1577382
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3575
- Ceres seq_id 1577383
- Location of start within SEQ ID NO 3574: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1596
- gi No. 5230728
- Description: (AF089851) peroxisomal copper-containing amine oxidase [Glycine max]
- % Identity: 72
- Alignment Length: 50
- Location of Alignment in SEQ ID NO 3575: from 4 to 49

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3576
- Ceres seq_id 1577384
- Location of start within SEQ ID NO 3574: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1597
- gi No. 5230728
- Description: (AF089851) peroxisomal copper-containing amine oxidase [Glycine max]
- % Identity: 72

- Alignment Length: 50
- Location of Alignment in SEQ ID NO 3576: from 1 to 30

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3577
- Ceres seq_id 1577385
- Location of start within SEQ ID NO 3574: at 443 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
277095
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3578
- Ceres seq_id 1577401
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3579
- Ceres seq_id 1577402
- Location of start within SEQ ID NO 3578: at 142 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1598
- gi No. 417298
- Description: MFS18 PROTEIN PRECURSOR >gi|100872|pir||S25103 MFS18 protein - maize >gi|22647|emb|CAA47738| (X67324) MFS18 [Zea mays]
- % Identity: 72.2
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 3579: from 64 to 142

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3580
- Ceres seq_id 1577403
- Location of start within SEQ ID NO 3578: at 148 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1599
- gi No. 417298
- Description: MFS18 PROTEIN PRECURSOR >gi|100872|pir||S25103 MFS18 protein - maize >gi|22647|emb|CAA47738| (X67324) MFS18 [Zea mays]
- % Identity: 72.2
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 3580: from 62 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3581
- Ceres seq_id 1577404
- Location of start within SEQ ID NO 3578: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1600
- gi No. 417298

- Description: MFS18 PROTEIN PRECURSOR >gi|100872|pir||S25103 MFS18 protein - maize >gi|22647|emb|CAA47738| (X67324) MFS18 [Zea mays]
- % Identity: 72.2
- Alignment Length: 79
- Location of Alignment in SEQ ID NO 3581: from 57 to 135

Maximum Length Sequence:
related to:
Clone IDs:
277121
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3582
- Ceres seq_id 1577405
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3583
- Ceres seq_id 1577406
- Location of start within SEQ ID NO 3582: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adaptor complexes medium subunit family
- Location within SEQ ID NO 3583: from 85 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1601
- gi No. 4874275
- Description: (AC007354) Similar to gb|L26291 clathrin-associated protein unc-101 from Caenorhabditis elegans and is a member of the PF|00928 Adapter complexes medium subunit family. [Arabidopsis thaliana]
- % Identity: 73.2
- Alignment Length: 82
- Location of Alignment in SEQ ID NO 3583: from 78 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3584
- Ceres seq_id 1577407
- Location of start within SEQ ID NO 3582: at 310 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adaptor complexes medium subunit family
- Location within SEQ ID NO 3584: from 58 to 175 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1602
- gi No. 4874275
- Description: (AC007354) Similar to gb|L26291 clathrin-associated protein unc-101 from Caenorhabditis elegans and is a member of the PF|00928 Adapter complexes medium subunit family. [Arabidopsis thaliana]
- % Identity: 91.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3584: from 58 to 175

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3585
- Ceres seq_id 1577408
- Location of start within SEQ ID NO 3582: at 436 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Adaptor complexes medium subunit family

- Location within SEQ ID NO 3585: from 16 to 133 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1603
- gi No. 4874275
- Description: (AC007354) Similar to gb|L26291 clathrin-associated protein unc-101 from Caenorhabditis elegans and is a member of the PF|00928 Adapter complexes medium subunit family. [Arabidopsis thaliana]
- % Identity: 91.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3585: from 16 to 133

Maximum Length Sequence:
related to:
Clone IDs:
277477
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3586
- Ceres seq_id 1577456
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3587
- Ceres seq_id 1577457
- Location of start within SEQ ID NO 3586: at 196 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3587: from 46 to 167 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1604
- gi No. 104563
- Description: calsequestrin, fast skeletal muscle - chicken >gi|211497 (M58048) calsequestrin [Gallus gallus]
- % Identity: 70.6
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 3587: from 39 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3588
- Ceres seq_id 1577458
- Location of start within SEQ ID NO 3586: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3588: from 36 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1605
- gi No. 104563
- Description: calsequestrin, fast skeletal muscle - chicken >gi|211497 (M58048) calsequestrin [Gallus gallus]
- % Identity: 70.6
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 3588: from 29 to 62

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3589
- Ceres seq_id 1577459
- Location of start within SEQ ID NO 3586: at 250 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic DNA topoisomerase I
- Location within SEQ ID NO 3589: from 28 to 149 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1606
- gi No. 104563
- Description: calsequestrin, fast skeletal muscle - chicken >gi|211497 (M58048) calsequestrin [Gallus gallus]
- % Identity: 70.6
- Alignment Length: 34
- Location of Alignment in SEQ ID NO 3589: from 21 to 54

Maximum Length Sequence:
related to:
Clone IDs:
277647
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3590
- Ceres seq_id 1577480
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3591
- Ceres seq_id 1577481
- Location of start within SEQ ID NO 3590: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3592
- Ceres seq_id 1577482
- Location of start within SEQ ID NO 3590: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3593
- Ceres seq_id 1577483
- Location of start within SEQ ID NO 3590: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1607
- gi No. 1707980
- Description: GLUTAREDOXIN >gi|1084458|pir||S41736 glutaredoxin thioltransferase - Rice >gi|485953|emb|CAA54397| (X77150) glutaredoxin [Oryza sativa]
- % Identity: 72.7
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 3593: from 17 to 60

Maximum Length Sequence:
related to:
Clone IDs:
277716

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3594
- Ceres seq_id 1577488

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3595
- Ceres seq_id 1577489
- Location of start within SEQ ID NO 3594: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1608
- gi No. 2191127
- Description: (AF007269) A_IG002N01.1 gene product [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 177
- Location of Alignment in SEQ ID NO 3595: from 93 to 269

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3596
- Ceres seq_id 1577490
- Location of start within SEQ ID NO 3594: at 29 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1609
- gi No. 2191127
- Description: (AF007269) A_IG002N01.1 gene product [Arabidopsis thaliana]
- % Identity: 76.3
- Alignment Length: 177
- Location of Alignment in SEQ ID NO 3596: from 84 to 260

Maximum Length Sequence:
related to:
Clone IDs:
277898

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3597
- Ceres seq_id 1577513

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3598
- Ceres seq_id 1577514
- Location of start within SEQ ID NO 3597: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1610
- gi No. 5734709
- Description: (AC008075) Contains PF|00069 Eukaryotic protein kinase domain. [Arabidopsis thaliana]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3598: from 80 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3599
- Ceres seq_id 1577515

- Location of start within SEQ ID NO 3597: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1611
- gi No. 2431767
- Description: (U62751) acidic ribosomal protein P3a [Zea mays]
- % Identity: 87.7
- Alignment Length: 57
- Location of Alignment in SEQ ID NO 3599: from 23 to 79

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3600
- Ceres seq_id 1577516
- Location of start within SEQ ID NO 3597: at 240 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
278006
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3601
- Ceres seq_id 1577517

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3602
- Ceres seq_id 1577518
- Location of start within SEQ ID NO 3601: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 3602: from 8 to 193 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1612
- gi No. 1710780
- Description: 40S RIBOSOMAL PROTEIN S9 (S7) >gi|1321917|emb|CAA65433| (X96613) cytoplasmic ribosomal protein S7 [Podospora anserina]
- % Identity: 71.7
- Alignment Length: 187
- Location of Alignment in SEQ ID NO 3602: from 8 to 193

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3603
- Ceres seq_id 1577519
- Location of start within SEQ ID NO 3601: at 356 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 3603: from 1 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1613
- gi No. 1710780

- Description: 40S RIBOSOMAL PROTEIN S9 (S7) >gi|1321917|emb|CAA65433| (X96613) cytoplasmic ribosomal protein S7 [Podospora anserina]
- % Identity: 71.7
- Alignment Length: 187
- Location of Alignment in SEQ ID NO 3603: from 1 to 112

Maximum Length Sequence:
related to:
Clone IDs:
278026
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3604
- Ceres seq_id 1577526
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3605
- Ceres seq_id 1577527
- Location of start within SEQ ID NO 3604: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1614
- gi No. 465445
- Description: PROBABLE NUCLEAR ANTIGEN >gi|418708|pir||B45344 probable nuclear antigen - suid herpesvirus 1 (strain Kaplan) >gi|334072 (M34651) ORF-3 protein [Pseudorabies virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3605: from 2 to 12

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3606
- Ceres seq_id 1577528
- Location of start within SEQ ID NO 3604: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3607
- Ceres seq_id 1577529
- Location of start within SEQ ID NO 3604: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
278095
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3608
- Ceres seq_id 1577539
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3609
- Ceres seq_id 1577540
- Location of start within SEQ ID NO 3608: at 209 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 3609: from 2 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1615
- gi No. 1173286
- Description: 40S RIBOSOMAL PROTEIN S9 >gi|112274|pir||S21497 ribosomal protein S9 - rat >gi|57143|emb|CAA47013| (X66370) ribosomal protein S9 [Rattus norvegicus]
- % Identity: 70.1
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3609: from 2 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3610
- Ceres seq_id 1577541
- Location of start within SEQ ID NO 3608: at 284 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 3610: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1616
- gi No. 1173286
- Description: 40S RIBOSOMAL PROTEIN S9 >gi|112274|pir||S21497 ribosomal protein S9 - rat >gi|57143|emb|CAA47013| (X66370) ribosomal protein S9 [Rattus norvegicus]
- % Identity: 70.1
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3610: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3611
- Ceres seq_id 1577542
- Location of start within SEQ ID NO 3608: at 398 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 3611: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1617
- gi No. 1173286
- Description: 40S RIBOSOMAL PROTEIN S9 >gi|112274|pir||S21497 ribosomal protein S9 - rat >gi|57143|emb|CAA47013| (X66370) ribosomal protein S9 [Rattus norvegicus]
- % Identity: 70.1
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3611: from 1 to 75

Maximum Length Sequence:
related to:
Clone IDs:
278291
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3612

- Ceres seq_id 1577548

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3613
- Ceres seq_id 1577549
- Location of start within SEQ ID NO 3612: at 62 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1618
- gi No. 1351357
- Description: UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 14 KD PROTEIN (CR14) >gi|633681|emb|CAA55863| (X79276) ubiquinol--cytochrome c reductase [Solanum tuberosum]
- % Identity: 73.2
- Alignment Length: 123
- Location of Alignment in SEQ ID NO 3613: from 1 to 123

Maximum Length Sequence:
related to:
Clone IDs:
278312

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3614
- Ceres seq_id 1577550

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3615
- Ceres seq_id 1577551
- Location of start within SEQ ID NO 3614: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3616
- Ceres seq_id 1577552
- Location of start within SEQ ID NO 3614: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1619
- gi No. 2668742
- Description: (AF034945) glycine-rich RNA binding protein [Zea mays]
- % Identity: 100
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3616: from 1 to 13

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3617
- Ceres seq_id 1577553
- Location of start within SEQ ID NO 3614: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1620
- gi No. 4038471

- Description: (AF111029) 40S ribosomal protein S27 homolog [Zea mays]
- % Identity: 100
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 3617: from 1 to 86

Maximum Length Sequence:
related to:
Clone IDs:
278781
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3618
- Ceres seq_id 1577582
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3619
- Ceres seq_id 1577583
- Location of start within SEQ ID NO 3618: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Herpesvirus processing and transport protein
- Location within SEQ ID NO 3619: from 31 to 158 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3620
- Ceres seq_id 1577584
- Location of start within SEQ ID NO 3618: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Herpesvirus processing and transport protein
- Location within SEQ ID NO 3620: from 18 to 145 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3621
- Ceres seq_id 1577585
- Location of start within SEQ ID NO 3618: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1621
- gi No. 2204081
- Description: (D86074) insoluble protein [Pinctada fucata]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3621: from 44 to 54

Maximum Length Sequence:
related to:
Clone IDs:
279070
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3622
- Ceres seq_id 1577601
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3623

- Ceres seq_id 1577602
- Location of start within SEQ ID NO 3622: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 3623: from 36 to 125 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3624
- Ceres seq_id 1577603
- Location of start within SEQ ID NO 3622: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 3624: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
279308
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3625
- Ceres seq_id 1577610

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3626
- Ceres seq_id 1577611
- Location of start within SEQ ID NO 3625: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1622
- gi No. 121631
- Description: GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR >gi|72323|pir||KNNT2S glycine-rich protein 2 - wood tobacco >gi|19743|emb|CAA42622| (X60007) nsGRP-2 [Nicotiana sylvestris]
- % Identity: 80.6
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 3626: from 97 to 129

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3627
- Ceres seq_id 1577612
- Location of start within SEQ ID NO 3625: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3628
- Ceres seq_id 1577613
- Location of start within SEQ ID NO 3625: at 232 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
279526
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3629
- Ceres seq_id 1577614
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3630
- Ceres seq_id 1577615
- Location of start within SEQ ID NO 3629: at 167 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 3630: from 18 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1623
- gi No. 5106775
- Description: (AF067732) ribosomal protein S12 [Hordeum vulgare]
- % Identity: 90.8
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 3630: from 1 to 139

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3631
- Ceres seq_id 1577616
- Location of start within SEQ ID NO 3629: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 3631: from 1 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1624
- gi No. 5106775
- Description: (AF067732) ribosomal protein S12 [Hordeum vulgare]
- % Identity: 90.8
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 3631: from 1 to 119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3632
- Ceres seq_id 1577617
- Location of start within SEQ ID NO 3629: at 236 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12e
- Location within SEQ ID NO 3632: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1625
- gi No. 5106775
- Description: (AF067732) ribosomal protein S12 [Hordeum vulgare]

- % Identity: 90.8
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 3632: from 1 to 116

Maximum Length Sequence:
related to:
Clone IDs:
279770
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3633
- Ceres seq_id 1577622
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3634
- Ceres seq_id 1577623
- Location of start within SEQ ID NO 3633: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3635
- Ceres seq_id 1577624
- Location of start within SEQ ID NO 3633: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3636
- Ceres seq_id 1577625
- Location of start within SEQ ID NO 3633: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1626
- gi No. 4494984
- Description: (AF083501) latent nuclear antigen [Macaca mulatta rhadinovirus 17577]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3636: from 13 to 23

Maximum Length Sequence:
related to:
Clone IDs:
280054
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3637
- Ceres seq_id 1577634
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3638
- Ceres seq_id 1577635
- Location of start within SEQ ID NO 3637: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3639
- Ceres seq_id 1577636
- Location of start within SEQ ID NO 3637: at 359 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1627
- gi No. 5103825
- Description: (AC007591) ESTs gb|AA650895, gb|AA720043 and gb|R29777 come from this gene. [Arabidopsis thaliana]
- % Identity: 88.7
- Alignment Length: 62
- Location of Alignment in SEQ ID NO 3639: from 2 to 62

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3640
- Ceres seq_id 1577637
- Location of start within SEQ ID NO 3637: at 544 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
280534
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3641
- Ceres seq_id 1577653

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3642
- Ceres seq_id 1577654
- Location of start within SEQ ID NO 3641: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3643
- Ceres seq_id 1577655
- Location of start within SEQ ID NO 3641: at 203 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1628
- gi No. 4512613
- Description: (AC004793) F28K20.12 [Arabidopsis thaliana]
- % Identity: 80.2
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 3643: from 1 to 82

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3644
- Ceres seq_id 1577656
- Location of start within SEQ ID NO 3641: at 361 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
280997
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3645
- Ceres seq_id 1577661
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3646
- Ceres seq_id 1577662
- Location of start within SEQ ID NO 3645: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S12
- Location within SEQ ID NO 3646: from 23 to 157 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1629
- gi No. 2829742
- Description: 40S RIBOSOMAL PROTEIN S23 >gi|1754684 (U81008) ribosomal protein S23 [Brugia malayi]
- % Identity: 75
- Alignment Length: 145
- Location of Alignment in SEQ ID NO 3646: from 15 to 158

Maximum Length Sequence:
related to:
Clone IDs:
281173
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3647
- Ceres seq_id 1577663
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3648
- Ceres seq_id 1577664
- Location of start within SEQ ID NO 3647: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3649
- Ceres seq_id 1577665
- Location of start within SEQ ID NO 3647: at 388 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L30
- Location within SEQ ID NO 3649: from 1 to 40 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1630
- gi No. 445613
- Description: ribosomal protein L7 [Solanum tuberosum]
- % Identity: 81.2

- Alignment Length: 223
- Location of Alignment in SEQ ID NO 3649: from 1 to 148

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3650
- Ceres seq_id 1577666
- Location of start within SEQ ID NO 3647: at 481 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1631
- gi No. 445613
- Description: ribosomal protein L7 [Solanum tuberosum]
- % Identity: 81.2
- Alignment Length: 223
- Location of Alignment in SEQ ID NO 3650: from 1 to 117

Maximum Length Sequence:
related to:
Clone IDs:
281533
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3651
- Ceres seq_id 1577678
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3652
- Ceres seq_id 1577679
- Location of start within SEQ ID NO 3651: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 3652: from 1 to 100 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3653
- Ceres seq_id 1577680
- Location of start within SEQ ID NO 3651: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 3653: from 1 to 69 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3654
- Ceres seq_id 1577681
- Location of start within SEQ ID NO 3651: at 103 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Aminotransferases class-I
- Location within SEQ ID NO 3654: from 1 to 66 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
281705
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3655
- Ceres seq_id 1577688
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3656
- Ceres seq_id 1577689
- Location of start within SEQ ID NO 3655: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 3656: from 195 to 254 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3657
- Ceres seq_id 1577690
- Location of start within SEQ ID NO 3655: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3658
- Ceres seq_id 1577691
- Location of start within SEQ ID NO 3655: at 196 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 3658: from 130 to 189 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
282040
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3659
- Ceres seq_id 1577698
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3660
- Ceres seq_id 1577699
- Location of start within SEQ ID NO 3659: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1632
- gi No. 4586449
- Description: (AB025187) cytochrome c oxidase subunit 6b-1 [Oryza sativa]
- % Identity: 83.9
- Alignment Length: 31

- Location of Alignment in SEQ ID NO 3660: from 50 to 80

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3661
- Ceres seq_id 1577700
- Location of start within SEQ ID NO 3659: at 518 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3662
- Ceres seq_id 1577701
- Location of start within SEQ ID NO 3659: at 557 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
282182
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3663
- Ceres seq_id 1577714
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3664
- Ceres seq_id 1577715
- Location of start within SEQ ID NO 3663: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3665
- Ceres seq_id 1577716
- Location of start within SEQ ID NO 3663: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 3665: from 55 to 133 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
282334
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3666
- Ceres seq_id 1577717
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3667
- Ceres seq_id 1577718
- Location of start within SEQ ID NO 3666: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1633
- gi No. 871830
- Description: (D30747) mini-collagen [Acropora donei]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3667: from 68 to 78

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3668
- Ceres seq_id 1577719
- Location of start within SEQ ID NO 3666: at 151 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
282468
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3669
- Ceres seq_id 1577737
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3670
- Ceres seq_id 1577738
- Location of start within SEQ ID NO 3669: at 141 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1634
- gi No. 82087
- Description: glycine-rich protein - tomato (fragment) >gi|19322|emb|CAA42538| (X59883) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 70
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3670: from 40 to 58

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3671
- Ceres seq_id 1577739
- Location of start within SEQ ID NO 3669: at 262 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1635
- gi No. 112994
- Description: GLYCINE-RICH RNA-BINDING, ABSCISIC ACID-INDUCIBLE PROTEIN >gi|82685|pir||S04536 embryonic abundant protein, glycine-rich - maize >gi|22313|emb|CAA31077| (X12564) ABA-inducible gene protein [Zea mays] >gi|226091|prf||1410284A
- % Identity: 76.9
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3671: from 98 to 110

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3672
- Ceres seq_id 1577740
- Location of start within SEQ ID NO 3669: at 277 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1636
- gi No. 112994
- Description: GLYCINE-RICH RNA-BINDING, ABSCISIC ACID-INDUCIBLE PROTEIN >gi|82685|pir||S04536 embryonic abundant protein, glycine-rich - maize >gi|22313|emb|CAA31077| (X12564) ABA-inducible gene protein [Zea mays] >gi|226091|prf||1410284A
- % Identity: 76.9
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3672: from 93 to 105

Maximum Length Sequence:
related to:
Clone IDs:
282494
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3673
- Ceres seq_id 1577749
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3674
- Ceres seq_id 1577750
- Location of start within SEQ ID NO 3673: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1637
- gi No. 688080
- Description: lectin=chitin-binding protein [Solanum tuberosum=potatoes, tubers, Peptide Partial, 27 aa]
- % Identity: 76.9
- Alignment Length: 13
- Location of Alignment in SEQ ID NO 3674: from 39 to 51

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3675
- Ceres seq_id 1577751
- Location of start within SEQ ID NO 3673: at 86 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3676
- Ceres seq_id 1577752
- Location of start within SEQ ID NO 3673: at 380 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:

Clone IDs:
282772
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3677
- Ceres seq_id 1577771
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3678
- Ceres seq_id 1577772
- Location of start within SEQ ID NO 3677: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3679
- Ceres seq_id 1577773
- Location of start within SEQ ID NO 3677: at 495 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1638
- gi No. 5123561
- Description: (AL079344) Thioredoxin-like protein [Arabidopsis thaliana]
- % Identity: 83.1
- Alignment Length: 59
- Location of Alignment in SEQ ID NO 3679: from 1 to 35

Maximum Length Sequence:
related to:
Clone IDs:
282824
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3680
- Ceres seq_id 1577774
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3681
- Ceres seq_id 1577775
- Location of start within SEQ ID NO 3680: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3682
- Ceres seq_id 1577776
- Location of start within SEQ ID NO 3680: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome c
- Location within SEQ ID NO 3682: from 83 to 173 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

282987
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3683
- Ceres seq_id 1577787
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3684
- Ceres seq_id 1577788
- Location of start within SEQ ID NO 3683: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3685
- Ceres seq_id 1577789
- Location of start within SEQ ID NO 3683: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 3685: from 88 to 116 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1639
- gi No. 729051
- Description: CALTRACTIN (CENTRIN) >gi|444342|prf||1906390A caltractin-like protein [Atriplex nummularia]
- % Identity: 70.9
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 3685: from 72 to 227

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3686
- Ceres seq_id 1577790
- Location of start within SEQ ID NO 3683: at 158 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 3686: from 36 to 64 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1640
- gi No. 729051
- Description: CALTRACTIN (CENTRIN) >gi|444342|prf||1906390A caltractin-like protein [Atriplex nummularia]
- % Identity: 70.9
- Alignment Length: 158
- Location of Alignment in SEQ ID NO 3686: from 20 to 175

Maximum Length Sequence:
related to:
Clone IDs:
283219
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3687
- Ceres seq_id 1577807
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3688
- Ceres seq_id 1577808

- Location of start within SEQ ID NO 3687: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1641
- gi No. 1314631
- Description: (U53220) pRb2/p130 [Homo sapiens]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3688: from 39 to 49

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3689
- Ceres seq_id 1577809
- Location of start within SEQ ID NO 3687: at 85 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3690
- Ceres seq_id 1577810
- Location of start within SEQ ID NO 3687: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
283380
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3691
- Ceres seq_id 1577813
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3692
- Ceres seq_id 1577814
- Location of start within SEQ ID NO 3691: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3693
- Ceres seq_id 1577815
- Location of start within SEQ ID NO 3691: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1642
- gi No. 4768968
- Description: (AF140219) nuclear cap-binding protein; CBP20 [Arabidopsis thaliana]
- % Identity: 77.6
- Alignment Length: 49
- Location of Alignment in SEQ ID NO 3693: from 27 to 75

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3694
- Ceres seq_id 1577816
- Location of start within SEQ ID NO 3691: at 245 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
283398
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3695
- Ceres seq_id 1577817

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3696
- Ceres seq_id 1577818
- Location of start within SEQ ID NO 3695: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3696: from 87 to 175 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3697
- Ceres seq_id 1577819
- Location of start within SEQ ID NO 3695: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3697: from 45 to 133 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3698
- Ceres seq_id 1577820
- Location of start within SEQ ID NO 3695: at 132 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3698: from 44 to 132 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
284027
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3699
- Ceres seq_id 1577845

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3700
- Ceres seq_id 1577846
- Location of start within SEQ ID NO 3699: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1643
- gi No. 401161
- Description: TUBULIN ALPHA-5 CHAIN >gi|322879|pir||S28982 tubulin alpha-5 chain - maize >gi|22156|emb|CAA44862| (X63177) alpha-tubulin #5 [Zea mays] >gi|450293 (L27815) alpha-tubulin [Zea mays] >gi|452474 (U05258) alpha-tubulin [Zea mays]
- % Identity: 100
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 3700: from 29 to 64

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3701
- Ceres seq_id 1577847
- Location of start within SEQ ID NO 3699: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
284289

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3702
- Ceres seq_id 1577866

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3703
- Ceres seq_id 1577867
- Location of start within SEQ ID NO 3702: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Profilins
- Location within SEQ ID NO 3703: from 50 to 174 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1644
- gi No. 3914423
- Description: PROFILIN 4 >gi|2642324 (AF032370) profilin [Zea mays]
- % Identity: 96.9
- Alignment Length: 131
- Location of Alignment in SEQ ID NO 3703: from 49 to 179

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3704
- Ceres seq_id 1577868
- Location of start within SEQ ID NO 3702: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Profilins
- Location within SEQ ID NO 3704: from 2 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1645
- gi No. 3914423
- Description: PROFILIN 4 >gi|2642324 (AF032370) profilin [Zea mays]
- % Identity: 96.9
- Alignment Length: 131
- Location of Alignment in SEQ ID NO 3704: from 1 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3705
- Ceres seq_id 1577869
- Location of start within SEQ ID NO 3702: at 180 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3705: from 1 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1646
- gi No. 3914423
- Description: PROFILIN 4 >gi|2642324 (AF032370) profilin [Zea mays]
- % Identity: 96.9
- Alignment Length: 131
- Location of Alignment in SEQ ID NO 3705: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
284988
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3706
- Ceres seq_id 1577881
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3707
- Ceres seq_id 1577882
- Location of start within SEQ ID NO 3706: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3707: from 22 to 70 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3708
- Ceres seq_id 1577883
- Location of start within SEQ ID NO 3706: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1647
- gi No. 2058273
- Description: (D83527) YK426 [Oryza sativa]
- % Identity: 86.7
- Alignment Length: 30

- Location of Alignment in SEQ ID NO 3708: from 1 to 30

Maximum Length Sequence:
related to:
Clone IDs:
285495
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3709
- Ceres seq_id 1577905
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3710
- Ceres seq_id 1577906
- Location of start within SEQ ID NO 3709: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 3710: from 47 to 78 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3711
- Ceres seq_id 1577907
- Location of start within SEQ ID NO 3709: at 56 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 3711: from 29 to 60 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3712
- Ceres seq_id 1577908
- Location of start within SEQ ID NO 3709: at 236 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
286407
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3713
- Ceres seq_id 1577940
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3714
- Ceres seq_id 1577941
- Location of start within SEQ ID NO 3713: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3715
- Ceres seq_id 1577942

- Location of start within SEQ ID NO 3713: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3716
- Ceres seq_id 1577943
- Location of start within SEQ ID NO 3713: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1648
- gi No. 123697
- Description: SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE) >gi|70787|pir||HSSH sperm histone - sheep
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3716: from 2 to 12

Maximum Length Sequence:
related to:
Clone IDs:
287015
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3717
- Ceres seq_id 1577969
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3718
- Ceres seq_id 1577970
- Location of start within SEQ ID NO 3717: at 14 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3718: from 67 to 134 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1649
- gi No. 118171
- Description: CYSTEINE PROTEINASE INHIBITOR-II (ORYZACYSTATIN-II) >gi|100692|pir||A38375 oryzacystatin II - rice >gi|169803 (J05595) oryzacystatin-II [Oryza sativa]
- % Identity: 75
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 3718: from 67 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3719
- Ceres seq_id 1577971
- Location of start within SEQ ID NO 3717: at 158 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3719: from 19 to 86 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1650

- gi No. 118171
- Description: CYSTEINE PROTEINASE INHIBITOR-II (ORYZACYSTATIN-II) >gi|100692|pir||A38375 oryzacystatin II - rice >gi|169803 (J05595) oryzacystatin-II [Oryza sativa]
- % Identity: 75
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 3719: from 19 to 86

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3720
- Ceres seq_id 1577972
- Location of start within SEQ ID NO 3717: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3720: from 9 to 76 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1651
- gi No. 118171
- Description: CYSTEINE PROTEINASE INHIBITOR-II (ORYZACYSTATIN-II) >gi|100692|pir||A38375 oryzacystatin II - rice >gi|169803 (J05595) oryzacystatin-II [Oryza sativa]
- % Identity: 75
- Alignment Length: 68
- Location of Alignment in SEQ ID NO 3720: from 9 to 76

Maximum Length Sequence:
related to:
Clone IDs:
287423
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3721
- Ceres seq_id 1577981

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3722
- Ceres seq_id 1577982
- Location of start within SEQ ID NO 3721: at 237 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 3722: from 11 to 129 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3723
- Ceres seq_id 1577983
- Location of start within SEQ ID NO 3721: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 3723: from 1 to 117 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3724

- Ceres seq_id 1577984
- Location of start within SEQ ID NO 3721: at 321 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L14
- Location within SEQ ID NO 3724: from 1 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
287498
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3725
- Ceres seq_id 1577999
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3726
- Ceres seq_id 1578000
- Location of start within SEQ ID NO 3725: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 3726: from 47 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1652
- gi No. 1498064
- Description: (U64825) AtE1 [Arabidopsis thaliana]
- % Identity: 71.6
- Alignment Length: 102
- Location of Alignment in SEQ ID NO 3726: from 103 to 198

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3727
- Ceres seq_id 1578001
- Location of start within SEQ ID NO 3725: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- Location within SEQ ID NO 3727: from 10 to 81 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1653
- gi No. 1498064
- Description: (U64825) AtE1 [Arabidopsis thaliana]
- % Identity: 71.6
- Alignment Length: 102
- Location of Alignment in SEQ ID NO 3727: from 66 to 161

Maximum Length Sequence:
related to:
Clone IDs:
287514
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3728
- Ceres seq_id 1578004
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3729
- Ceres seq_id 1578005
- Location of start within SEQ ID NO 3728: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AhpC/TSA family
- Location within SEQ ID NO 3729: from 88 to 239 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1654
- gi No. 3328221
- Description: (AF076920) thioredoxin peroxidase [Secale cereale]
- % Identity: 80.8
- Alignment Length: 261
- Location of Alignment in SEQ ID NO 3729: from 20 to 279

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3730
- Ceres seq_id 1578006
- Location of start within SEQ ID NO 3728: at 58 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- AhpC/TSA family
- Location within SEQ ID NO 3730: from 69 to 220 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1655
- gi No. 3328221
- Description: (AF076920) thioredoxin peroxidase [Secale cereale]
- % Identity: 80.8
- Alignment Length: 261
- Location of Alignment in SEQ ID NO 3730: from 1 to 260

Maximum Length Sequence:
related to:
Clone IDs:
287536
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3731
- Ceres seq_id 1578007

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3732
- Ceres seq_id 1578008
- Location of start within SEQ ID NO 3731: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3733
- Ceres seq_id 1578009
- Location of start within SEQ ID NO 3731: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3733: from 9 to 84 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3734
- Ceres seq_id 1578010
- Location of start within SEQ ID NO 3731: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
287563
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3735
- Ceres seq_id 1578015

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3736
- Ceres seq_id 1578016
- Location of start within SEQ ID NO 3735: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 3736: from 10 to 101 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3737
- Ceres seq_id 1578017
- Location of start within SEQ ID NO 3735: at 109 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 3737: from 1 to 65 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3738
- Ceres seq_id 1578018
- Location of start within SEQ ID NO 3735: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome P450
- Location within SEQ ID NO 3738: from 1 to 62 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
287654
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3739
- Ceres seq_id 1578031

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3740
- Ceres seq_id 1578032
- Location of start within SEQ ID NO 3739: at 140 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1656
- gi No. 1075072
- Description: nitrogen fixation protein (nifU) homolog - Haemophilus influenzae (strain Rd KW20)
- % Identity: 71.2
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 3740: from 34 to 158

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3741
- Ceres seq_id 1578033
- Location of start within SEQ ID NO 3739: at 143 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1657
- gi No. 1075072
- Description: nitrogen fixation protein (nifU) homolog - Haemophilus influenzae (strain Rd KW20)
- % Identity: 71.2
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 3741: from 33 to 157

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3742
- Ceres seq_id 1578034
- Location of start within SEQ ID NO 3739: at 353 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1658
- gi No. 1075072
- Description: nitrogen fixation protein (nifU) homolog - Haemophilus influenzae (strain Rd KW20)
- % Identity: 71.2
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 3742: from 1 to 87

Maximum Length Sequence:
related to:
Clone IDs:
287687
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3743
- Ceres seq_id 1578035
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3744
- Ceres seq_id 1578036
- Location of start within SEQ ID NO 3743: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Heme-binding domain in cytochrome b5 and oxidoreductases
- Location within SEQ ID NO 3744: from 7 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1659
- gi No. 729252
- Description: CYTOCHROME B5 >gi|167140 (M87514) cytochrome b-5 [Brassica oleracea] >gi|384338|prf||1905426A cytochrome b5 [Brassica oleracea]
- % Identity: 72.4
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 3744: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3745
- Ceres seq_id 1578037
- Location of start within SEQ ID NO 3743: at 294 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1660
- gi No. 729252
- Description: CYTOCHROME B5 >gi|167140 (M87514) cytochrome b-5 [Brassica oleracea] >gi|384338|prf||1905426A cytochrome b5 [Brassica oleracea]
- % Identity: 72.4
- Alignment Length: 134
- Location of Alignment in SEQ ID NO 3745: from 1 to 98

Maximum Length Sequence:
related to:
Clone IDs:
287883

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3746
- Ceres seq_id 1578042

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3747
- Ceres seq_id 1578043
- Location of start within SEQ ID NO 3746: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 3747: from 108 to 162 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3748
- Ceres seq_id 1578044
- Location of start within SEQ ID NO 3746: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3749
- Ceres seq_id 1578045
- Location of start within SEQ ID NO 3746: at 123 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
287904

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3750
- Ceres seq_id 1578056

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3751
- Ceres seq_id 1578057
- Location of start within SEQ ID NO 3750: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3752
- Ceres seq_id 1578058
- Location of start within SEQ ID NO 3750: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3752: from 40 to 87 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3753
- Ceres seq_id 1578059
- Location of start within SEQ ID NO 3750: at 107 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 3753: from 5 to 52 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
287981

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3754
- Ceres seq_id 1578070

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3755
- Ceres seq_id 1578071
- Location of start within SEQ ID NO 3754: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1661
- gi No. 5305335
- Description: (AF071081) proline-rich mucin homolog [Mycobacterium tuberculosis]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3755: from 34 to 44

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3756
- Ceres seq_id 1578073
- Location of start within SEQ ID NO 3754: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
288157
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3757
- Ceres seq_id 1578083
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3758
- Ceres seq_id 1578084
- Location of start within SEQ ID NO 3757: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic-type carbonic anhydrase
- Location within SEQ ID NO 3758: from 2 to 116 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3759
- Ceres seq_id 1578085
- Location of start within SEQ ID NO 3757: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic-type carbonic anhydrase
- Location within SEQ ID NO 3759: from 1 to 97 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
288179
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3760
- Ceres seq_id 1578086
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3761
- Ceres seq_id 1578087

- Location of start within SEQ ID NO 3760: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3762
- Ceres seq_id 1578088
- Location of start within SEQ ID NO 3760: at 449 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1662
- gi No. 1103318
- Description: (X78818) casein kinase I [Arabidopsis thaliana] >gi|2244791|emb|CAB10213.1| (Z97336) casein kinase I [Arabidopsis thaliana]
- % Identity: 79.5
- Alignment Length: 39
- Location of Alignment in SEQ ID NO 3762: from 1 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3763
- Ceres seq_id 1578089
- Location of start within SEQ ID NO 3760: at 533 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1663
- gi No. 1103318
- Description: (X78818) casein kinase I [Arabidopsis thaliana] >gi|2244791|emb|CAB10213.1| (Z97336) casein kinase I [Arabidopsis thaliana]
- % Identity: 79.5
- Alignment Length: 39
- Location of Alignment in SEQ ID NO 3763: from 1 to 11

Maximum Length Sequence:
related to:
Clone IDs:
288188
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3764
- Ceres seq_id 1578090
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3765
- Ceres seq_id 1578091
- Location of start within SEQ ID NO 3764: at 97 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Universal stress protein family
- Location within SEQ ID NO 3765: from 94 to 154 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
288889

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3766
- Ceres seq_id 1578130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3767
- Ceres seq_id 1578131
- Location of start within SEQ ID NO 3766: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3768
- Ceres seq_id 1578132
- Location of start within SEQ ID NO 3766: at 61 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1664
- gi No. 2341025
- Description: (AC000104) F19P19.2 [Arabidopsis thaliana]
- % Identity: 82
- Alignment Length: 61
- Location of Alignment in SEQ ID NO 3768: from 28 to 88

Maximum Length Sequence:
related to:
Clone IDs:
288996

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3769
- Ceres seq_id 1578150

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3770
- Ceres seq_id 1578151
- Location of start within SEQ ID NO 3769: at 265 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3771
- Ceres seq_id 1578152
- Location of start within SEQ ID NO 3769: at 492 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 3771: from 57 to 127 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3772
- Ceres seq_id 1578153
- Location of start within SEQ ID NO 3769: at 609 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- ATPases associated with various cellular activities (AAA)
- Location within SEQ ID NO 3772: from 18 to 88 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
289009
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3773
- Ceres seq_id 1578154
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3774
- Ceres seq_id 1578155
- Location of start within SEQ ID NO 3773: at 99 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 3774: from 71 to 98 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1665
- gi No. 3834317
- Description: (AC005679) Similar to CDC16 protein gb|U18291 (CDC16Hs) from Homo sapiens. [Arabidopsis thaliana]
- % Identity: 80.7
- Alignment Length: 114
- Location of Alignment in SEQ ID NO 3774: from 21 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3775
- Ceres seq_id 1578156
- Location of start within SEQ ID NO 3773: at 231 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- TPR Domain
- Location within SEQ ID NO 3775: from 27 to 54 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1666
- gi No. 3834317
- Description: (AC005679) Similar to CDC16 protein gb|U18291 (CDC16Hs) from Homo sapiens. [Arabidopsis thaliana]
- % Identity: 80.7
- Alignment Length: 114
- Location of Alignment in SEQ ID NO 3775: from 1 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3776
- Ceres seq_id 1578157
- Location of start within SEQ ID NO 3773: at 500 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1667

- gi No. 3834317
- Description: (AC005679) Similar to CDC16 protein gb|U18291 (CDC16Hs) from Homo sapiens. [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 3776: from 22 to 91

Maximum Length Sequence:
related to:
Clone IDs:
289218
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3777
- Ceres seq_id 1578185
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3778
- Ceres seq_id 1578186
- Location of start within SEQ ID NO 3777: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1668
- gi No. 3482918
- Description: (AC003970) Similar to ATP-citrate-lyase [Arabidopsis thaliana]
- % Identity: 89.6
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 3778: from 2 to 136

Maximum Length Sequence:
related to:
Clone IDs:
289240
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3779
- Ceres seq_id 1578190
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3780
- Ceres seq_id 1578191
- Location of start within SEQ ID NO 3779: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3781
- Ceres seq_id 1578192
- Location of start within SEQ ID NO 3779: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3782
- Ceres seq_id 1578193
- Location of start within SEQ ID NO 3779: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1669
- gi No. 5714684
- Description: (AF169203) differentially expressed osmotic protein ODE1 [Capsicum annuum]
- % Identity: 70.6
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 3782: from 97 to 113

Maximum Length Sequence:
related to:
Clone IDs:
289403
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3783
- Ceres seq_id 1578201
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3784
- Ceres seq_id 1578202
- Location of start within SEQ ID NO 3783: at 218 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1670
- gi No. 2760334
- Description: (AC002130) F1N21.5 [Arabidopsis thaliana]
- % Identity: 78.3
- Alignment Length: 23
- Location of Alignment in SEQ ID NO 3784: from 1 to 23

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3785
- Ceres seq_id 1578203
- Location of start within SEQ ID NO 3783: at 315 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3786
- Ceres seq_id 1578204
- Location of start within SEQ ID NO 3783: at 601 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
289813
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3787
- Ceres seq_id 1578239
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3788
- Ceres seq_id 1578240

- Location of start within SEQ ID NO 3787: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1671
- gi No. 551483
- Description: (U09276) ABA- and ripening-inducible-like protein [Zea mays]
- % Identity: 73.7
- Alignment Length: 133
- Location of Alignment in SEQ ID NO 3788: from 62 to 194

Maximum Length Sequence:
related to:
Clone IDs:
289947
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3789
- Ceres seq_id 1578247
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3790
- Ceres seq_id 1578248
- Location of start within SEQ ID NO 3789: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 3790: from 105 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1672
- gi No. 5733869
- Description: (AC007932) Contains PF|00646 F-box domain. [Arabidopsis thaliana]
- % Identity: 75
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3790: from 148 to 167

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3791
- Ceres seq_id 1578249
- Location of start within SEQ ID NO 3789: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1673
- gi No. 92559
- Description: nucleolin - rat
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3791: from 101 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3792
- Ceres seq_id 1578250
- Location of start within SEQ ID NO 3789: at 245 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1674
- gi No. 92559
- Description: nucleolin - rat
- % Identity: 72.2
- Alignment Length: 18
- Location of Alignment in SEQ ID NO 3792: from 95 to 112

Maximum Length Sequence:
related to:
Clone IDs:
289983
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3793
- Ceres seq_id 1578255
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3794
- Ceres seq_id 1578256
- Location of start within SEQ ID NO 3793: at 111 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1675
- gi No. 4572679
- Description: (AC006954) RSZp22 splicing factor; contains RNA recognition motif [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3794: from 157 to 170

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3795
- Ceres seq_id 1578257
- Location of start within SEQ ID NO 3793: at 258 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1676
- gi No. 4572679
- Description: (AC006954) RSZp22 splicing factor; contains RNA recognition motif [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3795: from 108 to 121

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3796
- Ceres seq_id 1578258
- Location of start within SEQ ID NO 3793: at 273 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1677
- gi No. 4572679
- Description: (AC006954) RSZp22 splicing factor; contains RNA recognition motif [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 14

- Location of Alignment in SEQ ID NO 3796: from 103 to 116

Maximum Length Sequence:
related to:
Clone IDs:
290563
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3797
- Ceres seq_id 1578317
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3798
- Ceres seq_id 1578318
- Location of start within SEQ ID NO 3797: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3799
- Ceres seq_id 1578319
- Location of start within SEQ ID NO 3797: at 179 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 3799: from 8 to 72 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1678
- gi No. 134126
- Description: SMALL NUCLEAR RIBONUCLEOPROTEIN E HOMOLOG C29 >gi|19573|emb|CAA44975| (X63376) snRNP-related protein [Medicago sativa]
- % Identity: 82.9
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 3799: from 1 to 76

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3800
- Ceres seq_id 1578320
- Location of start within SEQ ID NO 3797: at 218 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 3800: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1679
- gi No. 134126
- Description: SMALL NUCLEAR RIBONUCLEOPROTEIN E HOMOLOG C29 >gi|19573|emb|CAA44975| (X63376) snRNP-related protein [Medicago sativa]
- % Identity: 82.9
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 3800: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
290626
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3801
- Ceres seq_id 1578333

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3802
- Ceres seq_id 1578334
- Location of start within SEQ ID NO 3801: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Diacylglycerol kinase catalytic domain (presumed)
- Location within SEQ ID NO 3802: from 45 to 99 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
290630

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3803
- Ceres seq_id 1578335

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3804
- Ceres seq_id 1578336
- Location of start within SEQ ID NO 3803: at 74 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Plant lipid transfer protein family
- Location within SEQ ID NO 3804: from 34 to 122 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
290806

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3805
- Ceres seq_id 1578352

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3806
- Ceres seq_id 1578353
- Location of start within SEQ ID NO 3805: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1680
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 3806: from 44 to 173

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3807
- Ceres seq_id 1578354
- Location of start within SEQ ID NO 3805: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1681
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 3807: from 21 to 150

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3808
- Ceres seq_id 1578355
- Location of start within SEQ ID NO 3805: at 244 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1682
- gi No. 4490705
- Description: (AL035680) ribosomal protein L14-like protein [Arabidopsis thaliana]
- % Identity: 81.5
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 3808: from 1 to 92

Maximum Length Sequence:
related to:
Clone IDs:
291213

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3809
- Ceres seq_id 1578379

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3810
- Ceres seq_id 1578380
- Location of start within SEQ ID NO 3809: at 88 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1683
- gi No. 228938
- Description: Hyp-rich glycoprotein [Zea diploperennis]
- % Identity: 71.8
- Alignment Length: 44
- Location of Alignment in SEQ ID NO 3810: from 7 to 44

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3811
- Ceres seq_id 1578381
- Location of start within SEQ ID NO 3809: at 226 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3812
- Ceres seq_id 1578382

- Location of start within SEQ ID NO 3809: at 278 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
291272
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3813
- Ceres seq_id 1578386
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3814
- Ceres seq_id 1578387
- Location of start within SEQ ID NO 3813: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3814: from 56 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1684
- gi No. 3176108
- Description: (AB001869) Brain-2 gene [Anolis carolinensis]
- % Identity: 100
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3814: from 97 to 107

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3815
- Ceres seq_id 1578388
- Location of start within SEQ ID NO 3813: at 221 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3816
- Ceres seq_id 1578389
- Location of start within SEQ ID NO 3813: at 408 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
291335
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3817
- Ceres seq_id 1578394
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3818
- Ceres seq_id 1578395
- Location of start within SEQ ID NO 3817: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3819
- Ceres seq_id 1578396
- Location of start within SEQ ID NO 3817: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 3819: from 9 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1685
- gi No. 1532175
- Description: (U63815) similar to protein disulfide isomerase [Arabidopsis thaliana]
- % Identity: 79
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 3819: from 9 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3820
- Ceres seq_id 1578397
- Location of start within SEQ ID NO 3817: at 295 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thioredoxin
- Location within SEQ ID NO 3820: from 1 to 58 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1686
- gi No. 1532175
- Description: (U63815) similar to protein disulfide isomerase [Arabidopsis thaliana]
- % Identity: 79
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 3820: from 1 to 58

Maximum Length Sequence:
related to:
Clone IDs:
291613
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3821
- Ceres seq_id 1578416
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3822
- Ceres seq_id 1578417
- Location of start within SEQ ID NO 3821: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3823
- Ceres seq_id 1578418

- Location of start within SEQ ID NO 3821: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynein light chain type 1
- Location within SEQ ID NO 3823: from 60 to 144 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3824
- Ceres seq_id 1578419
- Location of start within SEQ ID NO 3821: at 50 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Dynein light chain type 1
- Location within SEQ ID NO 3824: from 44 to 128 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
291629
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3825
- Ceres seq_id 1578422
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3826
- Ceres seq_id 1578423
- Location of start within SEQ ID NO 3825: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3827
- Ceres seq_id 1578424
- Location of start within SEQ ID NO 3825: at 250 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic initiation factor 5A hypusine (eIF-5A)
- Location within SEQ ID NO 3827: from 1 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1687
- gi No. 5080771
- Description: (AC007576) initiation factor 5A-4 [Arabidopsis thaliana]
- % Identity: 76.6
- Alignment Length: 111
- Location of Alignment in SEQ ID NO 3827: from 1 to 97

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3828
- Ceres seq_id 1578425
- Location of start within SEQ ID NO 3825: at 594 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
291924
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3829
- Ceres seq_id 1578443
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3830
- Ceres seq_id 1578444
- Location of start within SEQ ID NO 3829: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3831
- Ceres seq_id 1578445
- Location of start within SEQ ID NO 3829: at 122 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1688
- gi No. 2493695
- Description: PHOTOSYSTEM II REACTION CENTRE W PROTEIN (PSII 6.1 KD PROTEIN)
- % Identity: 100
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3831: from 74 to 93

Maximum Length Sequence:
related to:
Clone IDs:
291949
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3832
- Ceres seq_id 1578448
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3833
- Ceres seq_id 1578449
- Location of start within SEQ ID NO 3832: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3834
- Ceres seq_id 1578450
- Location of start within SEQ ID NO 3832: at 396 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3834: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1689
- gi No. 4006890
- Description: (Z99708) ubiquitin--protein ligase-like protein [Arabidopsis thaliana]
- % Identity: 77.9
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 3834: from 1 to 124

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3835
- Ceres seq_id 1578451
- Location of start within SEQ ID NO 3832: at 438 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3835: from 1 to 99 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1690
- gi No. 4006890
- Description: (Z99708) ubiquitin--protein ligase-like protein [Arabidopsis thaliana]
- % Identity: 77.9
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 3835: from 1 to 110

Maximum Length Sequence:
related to:
Clone IDs:
291952
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3836
- Ceres seq_id 1578452
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3837
- Ceres seq_id 1578453
- Location of start within SEQ ID NO 3836: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3838
- Ceres seq_id 1578454
- Location of start within SEQ ID NO 3836: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1691
- gi No. 1351285
- Description: TROPOMYOSIN 1, FUSION PROTEIN 33 >gi|433474|emb|CAA53800| (X76208) protein 33-specific exons [Drosophila melanogaster]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3838: from 67 to 80

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3839
- Ceres seq_id 1578455
- Location of start within SEQ ID NO 3836: at 41 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1692
- gi No. 1351285
- Description: TROPOMYOSIN 1, FUSION PROTEIN 33 >gi|433474|emb|CAA53800| (X76208) protein 33-specific exons [Drosophila melanogaster]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3839: from 54 to 67

Maximum Length Sequence:
related to:
Clone IDs:
291984
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3840
- Ceres seq_id 1578464
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3841
- Ceres seq_id 1578465
- Location of start within SEQ ID NO 3840: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3842
- Ceres seq_id 1578466
- Location of start within SEQ ID NO 3840: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1693
- gi No. 2246378
- Description: (Z86094) plastid protein [Arabidopsis thaliana]
- % Identity: 72.8
- Alignment Length: 127
- Location of Alignment in SEQ ID NO 3842: from 67 to 187

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3843
- Ceres seq_id 1578467
- Location of start within SEQ ID NO 3840: at 72 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1694
- gi No. 2246378
- Description: (Z86094) plastid protein [Arabidopsis thaliana]
- % Identity: 72.8

- Alignment Length: 127
- Location of Alignment in SEQ ID NO 3843: from 44 to 164

Maximum Length Sequence:
related to:
Clone IDs:
291986
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3844
- Ceres seq_id 1578468
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3845
- Ceres seq_id 1578469
- Location of start within SEQ ID NO 3844: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3846
- Ceres seq_id 1578470
- Location of start within SEQ ID NO 3844: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3847
- Ceres seq_id 1578471
- Location of start within SEQ ID NO 3844: at 357 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1695
- gi No. 3850571
- Description: (AC005278) Similar to gb|U85207 snRNP core Sm protein homolog Sm-X5 from Mus musculus. EST gb|AA612141 comes from this gene. [Arabidopsis thaliana]
- % Identity: 91.9
- Alignment Length: 86
- Location of Alignment in SEQ ID NO 3847: from 1 to 40

Maximum Length Sequence:
related to:
Clone IDs:
292039
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3848
- Ceres seq_id 1578472
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3849
- Ceres seq_id 1578473
- Location of start within SEQ ID NO 3848: at 64 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Germin family
- Location within SEQ ID NO 3849: from 24 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1696
- gi No. 348652
- Description: 25K protein 4222 - rice (strain Nihonbare) (fragment)
- % Identity: 85
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3849: from 23 to 42

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3850
- Ceres seq_id 1578474
- Location of start within SEQ ID NO 3848: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Germin family
- Location within SEQ ID NO 3850: from 21 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1697
- gi No. 348652
- Description: 25K protein 4222 - rice (strain Nihonbare) (fragment)
- % Identity: 85
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3850: from 20 to 39

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3851
- Ceres seq_id 1578475
- Location of start within SEQ ID NO 3848: at 487 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
292093
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3852
- Ceres seq_id 1578479
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3853
- Ceres seq_id 1578480
- Location of start within SEQ ID NO 3852: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3853: from 7 to 63 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1698
- gi No. 4808585
- Description: (AF093748) KH type splicing regulatory protein; KSRP [Homo sapiens]
- % Identity: 72.7

- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3853: from 44 to 54

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3854
- Ceres seq_id 1578481
- Location of start within SEQ ID NO 3852: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3855
- Ceres seq_id 1578482
- Location of start within SEQ ID NO 3852: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
292207
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3856
- Ceres seq_id 1578486
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3857
- Ceres seq_id 1578487
- Location of start within SEQ ID NO 3856: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3857: from 68 to 160 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1699
- gi No. 399334
- Description: CYSTATIN I PRECURSOR (CORN KERNEL CYSTEINE PROTEINASE INHIBITOR) >gi|322868|pir||S27239 cysteine proteinase inhibitor - maize >gi|217962|dbj|BAA01472| (D10622) corn cystatin I [Zea mays]
- % Identity: 97.8
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 3857: from 27 to 161

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3858
- Ceres seq_id 1578488
- Location of start within SEQ ID NO 3856: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3858: from 42 to 134 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1700
- gi No. 399334

- Description: CYSTATIN I PRECURSOR (CORN KERNEL CYSTEINE PROTEINASE INHIBITOR) >gi|322868|pir||S27239 cysteine proteinase inhibitor - maize >gi|217962|dbj|BAA01472| (D10622) corn cystatin I [Zea mays]
- % Identity: 97.8
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 3858: from 1 to 135

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3859
- Ceres seq_id 1578489
- Location of start within SEQ ID NO 3856: at 195 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cystatin domain
- Location within SEQ ID NO 3859: from 4 to 96 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1701
- gi No. 399334
- Description: CYSTATIN I PRECURSOR (CORN KERNEL CYSTEINE PROTEINASE INHIBITOR) >gi|322868|pir||S27239 cysteine proteinase inhibitor - maize >gi|217962|dbj|BAA01472| (D10622) corn cystatin I [Zea mays]
- % Identity: 97.8
- Alignment Length: 135
- Location of Alignment in SEQ ID NO 3859: from 1 to 97

Maximum Length Sequence:
related to:
Clone IDs:
292229
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3860
- Ceres seq_id 1578490
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3861
- Ceres seq_id 1578491
- Location of start within SEQ ID NO 3860: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1702
- gi No. 2388584
- Description: (AC000098) EST gb|ATTS1136 comes from this gene. [Arabidopsis thaliana]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 3861: from 128 to 139

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3862
- Ceres seq_id 1578492
- Location of start within SEQ ID NO 3860: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3863

- Ceres seq_id 1578493
- Location of start within SEQ ID NO 3860: at 360 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
292445
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3864
- Ceres seq_id 1578510
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3865
- Ceres seq_id 1578511
- Location of start within SEQ ID NO 3864: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3866
- Ceres seq_id 1578512
- Location of start within SEQ ID NO 3864: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1703
- gi No. 5456996
- Description: (AB024380) ORF1 [TT virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3866: from 72 to 82

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3867
- Ceres seq_id 1578513
- Location of start within SEQ ID NO 3864: at 303 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
292727
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3868
- Ceres seq_id 1578524
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3869
- Ceres seq_id 1578525
- Location of start within SEQ ID NO 3868: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3870
- Ceres seq_id 1578526
- Location of start within SEQ ID NO 3868: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1704
- gi No. 3935169
- Description: (AC004557) F17L21.12 [Arabidopsis thaliana]
- % Identity: 76.1
- Alignment Length: 67
- Location of Alignment in SEQ ID NO 3870: from 67 to 132

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3871
- Ceres seq_id 1578527
- Location of start within SEQ ID NO 3868: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1705
- gi No. 3935169
- Description: (AC004557) F17L21.12 [Arabidopsis thaliana]
- % Identity: 76.1
- Alignment Length: 67
- Location of Alignment in SEQ ID NO 3871: from 1 to 66

Maximum Length Sequence:
related to:
Clone IDs:
292855

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3872
- Ceres seq_id 1578540

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3873
- Ceres seq_id 1578541
- Location of start within SEQ ID NO 3872: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 3873: from 17 to 54 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1706
- gi No. 127586
- Description: ANTHOCYANIN REGULATORY C1-I PROTEIN >gi|100858|pir||S12369 C1-I protein - maize >gi|22214|emb|CAA36456| (X52201) C1-I [Zea mays]
- % Identity: 77.3
- Alignment Length: 22
- Location of Alignment in SEQ ID NO 3873: from 33 to 54

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3874

- Ceres seq_id 1578542
- Location of start within SEQ ID NO 3872: at 35 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Myb-like DNA-binding domain
- Location within SEQ ID NO 3874: from 6 to 43 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1707
- gi No. 127586
- Description: ANTHOCYANIN REGULATORY C1-I PROTEIN >gi|100858|pir||S12369 C1-I protein - maize >gi|22214|emb|CAA36456| (X52201) C1-I [Zea mays]
- % Identity: 77.3
- Alignment Length: 22
- Location of Alignment in SEQ ID NO 3874: from 22 to 43

Maximum Length Sequence:
related to:
Clone IDs:
292861
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3875
- Ceres seq_id 1578543

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3876
- Ceres seq_id 1578544
- Location of start within SEQ ID NO 3875: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7Ae
- Location within SEQ ID NO 3876: from 74 to 172 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1708
- gi No. 4826860
- Description: ref|NP_004999.1|pNHP2L1| non-histone chromosome protein 2 (S. cerevisiae)-like 1 >gi|2500345|sp|P55769|NHPX_HUMAN NHP2/RS6 FAMILY PROTEIN YEL026W HOMOLOG (HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN 2 >gi|3859990 (AF091076) OTK27 [Homo sapiens]
- % Identity: 80.2
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3876: from 59 to 184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3877
- Ceres seq_id 1578545
- Location of start within SEQ ID NO 3875: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7Ae
- Location within SEQ ID NO 3877: from 17 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1709
- gi No. 4826860
- Description: ref|NP_004999.1|pNHP2L1| non-histone chromosome protein 2 (S. cerevisiae)-like 1 >gi|2500345|sp|P55769|NHPX_HUMAN NHP2/RS6

FAMILY PROTEIN YEL026W HOMOLOG (HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN 2
>gi|3859990 (AF091076) OTK27 [Homo sapiens]
- % Identity: 80.2
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3877: from 2 to 127

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3878
- Ceres seq_id 1578546
- Location of start within SEQ ID NO 3875: at 227 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7Ae
- Location within SEQ ID NO 3878: from 1 to 97 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1710
- gi No. 4826860
- Description: ref|NP_004999.1|pNHP2L1| non-histone chromosome protein 2 (S. cerevisiae)-like 1 >gi|2500345|sp|P55769|NHPX_HUMAN NHP2/RS6 FAMILY PROTEIN YEL026W HOMOLOG (HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN 2 >gi|3859990 (AF091076) OTK27 [Homo sapiens]
- % Identity: 80.2
- Alignment Length: 126
- Location of Alignment in SEQ ID NO 3878: from 1 to 109

Maximum Length Sequence:
related to:
Clone IDs:
293088
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3879
- Ceres seq_id 1578582
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3880
- Ceres seq_id 1578583
- Location of start within SEQ ID NO 3879: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3881
- Ceres seq_id 1578584
- Location of start within SEQ ID NO 3879: at 71 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1711
- gi No. 4769012
- Description: (AF143746) CER1 [Oryza sativa]
- % Identity: 95
- Alignment Length: 20
- Location of Alignment in SEQ ID NO 3881: from 1 to 20

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3882
- Ceres seq_id 1578585

- Location of start within SEQ ID NO 3879: at 247 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
293160
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3883
- Ceres seq_id 1578586
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3884
- Ceres seq_id 1578587
- Location of start within SEQ ID NO 3883: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Reverse transcriptase (RNA-dependent DNA polymerase)
- Location within SEQ ID NO 3884: from 21 to 79 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3885
- Ceres seq_id 1578588
- Location of start within SEQ ID NO 3883: at 266 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3886
- Ceres seq_id 1578589
- Location of start within SEQ ID NO 3883: at 269 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
293228
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3887
- Ceres seq_id 1578590
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3888
- Ceres seq_id 1578591
- Location of start within SEQ ID NO 3887: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3889

- Ceres seq_id 1578592
- Location of start within SEQ ID NO 3887: at 92 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oleosin
- Location within SEQ ID NO 3889: from 1 to 78 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3890
- Ceres seq_id 1578593
- Location of start within SEQ ID NO 3887: at 161 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Oleosin
- Location within SEQ ID NO 3890: from 1 to 55 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
293895
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3891
- Ceres seq_id 1578630
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3892
- Ceres seq_id 1578631
- Location of start within SEQ ID NO 3891: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 3892: from 23 to 65 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3893
- Ceres seq_id 1578632
- Location of start within SEQ ID NO 3891: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat
- Location within SEQ ID NO 3893: from 1 to 32 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3894
- Ceres seq_id 1578633
- Location of start within SEQ ID NO 3891: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- WD domain, G-beta repeat

- Location within SEQ ID NO 3894: from 1 to 31 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
294494
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3895
- Ceres seq_id 1578674
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3896
- Ceres seq_id 1578675
- Location of start within SEQ ID NO 3895: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- UBX domain
- Location within SEQ ID NO 3896: from 4 to 76 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3897
- Ceres seq_id 1578676
- Location of start within SEQ ID NO 3895: at 25 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- UBX domain
- Location within SEQ ID NO 3897: from 1 to 68 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3898
- Ceres seq_id 1578677
- Location of start within SEQ ID NO 3895: at 175 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
294577
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3899
- Ceres seq_id 1578682
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3900
- Ceres seq_id 1578683
- Location of start within SEQ ID NO 3899: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- metallopeptidase family M24
- Location within SEQ ID NO 3900: from 1 to 169 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1712
- gi No. 5123544
- Description: (AL079344) X-Pro dipeptidase-like protein (fragment) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 3900: from 7 to 184

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3901
- Ceres seq_id 1578684
- Location of start within SEQ ID NO 3899: at 55 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- metallopeptidase family M24
- Location within SEQ ID NO 3901: from 1 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1713
- gi No. 5123544
- Description: (AL079344) X-Pro dipeptidase-like protein (fragment) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 3901: from 1 to 166

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3902
- Ceres seq_id 1578685
- Location of start within SEQ ID NO 3899: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- metallopeptidase family M24
- Location within SEQ ID NO 3902: from 1 to 144 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1714
- gi No. 5123544
- Description: (AL079344) X-Pro dipeptidase-like protein (fragment) [Arabidopsis thaliana]
- % Identity: 72.5
- Alignment Length: 178
- Location of Alignment in SEQ ID NO 3902: from 1 to 159

Maximum Length Sequence:
related to:
Clone IDs:
294752
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3903
- Ceres seq_id 1578692

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3904
- Ceres seq_id 1578693
- Location of start within SEQ ID NO 3903: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3905
- Ceres seq_id 1578694
- Location of start within SEQ ID NO 3903: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L6
- Location within SEQ ID NO 3905: from 9 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
294798
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3906
- Ceres seq_id 1578695
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3907
- Ceres seq_id 1578696
- Location of start within SEQ ID NO 3906: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal S17
- Location within SEQ ID NO 3907: from 52 to 115 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1715
- gi No. 1350944
- Description: 40S RIBOSOMAL PROTEIN S17
- % Identity: 87.7
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 3907: from 51 to 115

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3908
- Ceres seq_id 1578697
- Location of start within SEQ ID NO 3906: at 153 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal S17
- Location within SEQ ID NO 3908: from 2 to 65 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1716
- gi No. 1350944
- Description: 40S RIBOSOMAL PROTEIN S17
- % Identity: 87.7
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 3908: from 1 to 65

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3909
- Ceres seq_id 1578698
- Location of start within SEQ ID NO 3906: at 222 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1717
- gi No. 1350944
- Description: 40S RIBOSOMAL PROTEIN S17
- % Identity: 87.7
- Alignment Length: 65
- Location of Alignment in SEQ ID NO 3909: from 1 to 42

Maximum Length Sequence:
related to:
Clone IDs:
294815
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3910
- Ceres seq_id 1578699
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3911
- Ceres seq_id 1578700
- Location of start within SEQ ID NO 3910: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3912
- Ceres seq_id 1578701
- Location of start within SEQ ID NO 3910: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1718
- gi No. 553165
- Description: (M76539) acetylcholinesterase [Homo sapiens]
- % Identity: 81.8
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3912: from 7 to 17

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3913
- Ceres seq_id 1578702
- Location of start within SEQ ID NO 3910: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
294826
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3914
- Ceres seq_id 1578703
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3915
- Ceres seq_id 1578704

- Location of start within SEQ ID NO 3914: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 3915: from 110 to 146 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3916
- Ceres seq_id 1578705
- Location of start within SEQ ID NO 3914: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 3916: from 87 to 123 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
295353
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3917
- Ceres seq_id 1578737
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3918
- Ceres seq_id 1578738
- Location of start within SEQ ID NO 3917: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1719
- gi No. 465445
- Description: PROBABLE NUCLEAR ANTIGEN >gi|418708|pir||B45344 probable nuclear antigen - suid herpesvirus 1 (strain Kaplan) >gi|334072 (M34651) ORF-3 protein [Pseudorabies virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3918: from 25 to 35

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3919
- Ceres seq_id 1578739
- Location of start within SEQ ID NO 3917: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome c oxidase subunit Vb
- Location within SEQ ID NO 3919: from 75 to 177 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1720
- gi No. 1841355
- Description: (D85381) cytochrome c oxidase subunit Vb precursor [Oryza sativa]
- % Identity: 85.8

- Alignment Length: 155
- Location of Alignment in SEQ ID NO 3919: from 27 to 177

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3920
- Ceres seq_id 1578740
- Location of start within SEQ ID NO 3917: at 80 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytochrome c oxidase subunit Vb
- Location within SEQ ID NO 3920: from 49 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1721
- gi No. 1841355
- Description: (D85381) cytochrome c oxidase subunit Vb precursor [Oryza sativa]
- % Identity: 85.8
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 3920: from 1 to 151

Maximum Length Sequence:
related to:
Clone IDs:
295421
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3921
- Ceres seq_id 1578744

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3922
- Ceres seq_id 1578745
- Location of start within SEQ ID NO 3921: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3923
- Ceres seq_id 1578746
- Location of start within SEQ ID NO 3921: at 425 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3923: from 1 to 81 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
295783
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3924
- Ceres seq_id 1578754

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3925
- Ceres seq_id 1578755
- Location of start within SEQ ID NO 3924: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3926
- Ceres seq_id 1578756
- Location of start within SEQ ID NO 3924: at 377 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1722
- gi No. 729535
- Description: FERREDOXIN-THIOREDOXIN REDUCTASE, CATALYTIC CHAIN PRECURSOR (FTR-C) (FERREDOXIN-THIOREDOXIN REDUCTASE SUBUNIT B) (FTR-B)
- % Identity: 89.6
- Alignment Length: 148
- Location of Alignment in SEQ ID NO 3926: from 1 to 103

Maximum Length Sequence:
related to:
Clone IDs:
295878
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3927
- Ceres seq_id 1578759
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3928
- Ceres seq_id 1578760
- Location of start within SEQ ID NO 3927: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1723
- gi No. 3123264
- Description: 60S RIBOSOMAL PROTEIN L27
>gi|2244857|emb|CAB10279.1| (Z97337) ribosomal protein [Arabidopsis thaliana]
- % Identity: 81.8
- Alignment Length: 137
- Location of Alignment in SEQ ID NO 3928: from 1 to 137

Maximum Length Sequence:
related to:
Clone IDs:
296071
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3929
- Ceres seq_id 1578777
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3930
- Ceres seq_id 1578778
- Location of start within SEQ ID NO 3929: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Histone deacetylase family
- Location within SEQ ID NO 3930: from 1 to 125 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3931
- Ceres seq_id 1578779
- Location of start within SEQ ID NO 3929: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
296191
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3932
- Ceres seq_id 1578782
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3933
- Ceres seq_id 1578783
- Location of start within SEQ ID NO 3932: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3934
- Ceres seq_id 1578785
- Location of start within SEQ ID NO 3932: at 340 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1724
- gi No. 3123515
- Description: (Y08761) Mago Nashi-like protein [Euphorbia lagascae]
- % Identity: 89
- Alignment Length: 145
- Location of Alignment in SEQ ID NO 3934: from 1 to 102

Maximum Length Sequence:
related to:
Clone IDs:
296214
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3935
- Ceres seq_id 1578786
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3936
- Ceres seq_id 1578787
- Location of start within SEQ ID NO 3935: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3937

- Ceres seq_id 1578788
- Location of start within SEQ ID NO 3935: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3938
- Ceres seq_id 1578789
- Location of start within SEQ ID NO 3935: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1725
- gi No. 1825645
- Description: (U88173) weak similarity to Arabidopsis thaliana ubiquitin-like protein 8 [Caenorhabditis elegans]
- % Identity: 75
- Alignment Length: 72
- Location of Alignment in SEQ ID NO 3938: from 1 to 72

Maximum Length Sequence:
related to:
Clone IDs:
296232
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3939
- Ceres seq_id 1578793
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3940
- Ceres seq_id 1578794
- Location of start within SEQ ID NO 3939: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Phosphopantetheine attachment site
- Location within SEQ ID NO 3940: from 84 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1726
- gi No. 113169
- Description: ACYL CARRIER PROTEIN II PRECURSOR (ACP II) >gi|166969 (M63799) acyl carrier protein II [Hordeum vulgare] >gi|228694|prf||1808324A acyl carrier protein II [Hordeum vulgare]
- % Identity: 81.3
- Alignment Length: 107
- Location of Alignment in SEQ ID NO 3940: from 49 to 155

Maximum Length Sequence:
related to:
Clone IDs:
296307
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3941
- Ceres seq_id 1578799
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3942
- Ceres seq_id 1578800
- Location of start within SEQ ID NO 3941: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3942: from 11 to 105 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3943
- Ceres seq_id 1578801
- Location of start within SEQ ID NO 3941: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3943: from 1 to 82 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3944
- Ceres seq_id 1578802
- Location of start within SEQ ID NO 3941: at 158 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
296630

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3945
- Ceres seq_id 1578817

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3946
- Ceres seq_id 1578818
- Location of start within SEQ ID NO 3945: at 229 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ubiquitin-conjugating enzyme
- Location within SEQ ID NO 3946: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1727
- gi No. 3915186
- Description: UBIQUITIN-CONJUGATING ENZYME E2-21 KD (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN) (PEROXIN-4) >gi|3128447 (AF061604) ubiquitin-conjugating enzyme homolog peroxin 4 [Pichia angusta]
- % Identity: 72.7
- Alignment Length: 22
- Location of Alignment in SEQ ID NO 3946: from 79 to 100

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3947
- Ceres seq_id 1578819
- Location of start within SEQ ID NO 3945: at 572 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
296765
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3948
- Ceres seq_id 1578834
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3949
- Ceres seq_id 1578835
- Location of start within SEQ ID NO 3948: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 3949: from 38 to 84 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3950
- Ceres seq_id 1578836
- Location of start within SEQ ID NO 3948: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Gamma-thionins family
- Location within SEQ ID NO 3950: from 31 to 77 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3951
- Ceres seq_id 1578837
- Location of start within SEQ ID NO 3948: at 243 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
296813
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3952
- Ceres seq_id 1578845
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3953
- Ceres seq_id 1578846
- Location of start within SEQ ID NO 3952: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S1 RNA binding domain
- Location within SEQ ID NO 3953: from 140 to 210 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3954
- Ceres seq_id 1578847
- Location of start within SEQ ID NO 3952: at 173 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S1 RNA binding domain
- Location within SEQ ID NO 3954: from 83 to 153 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3955
- Ceres seq_id 1578848
- Location of start within SEQ ID NO 3952: at 194 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- S1 RNA binding domain
- Location within SEQ ID NO 3955: from 76 to 146 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
296827
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3956
- Ceres seq_id 1578852

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3957
- Ceres seq_id 1578853
- Location of start within SEQ ID NO 3956: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain
- Location within SEQ ID NO 3957: from 123 to 181 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1728
- gi No. 322647
- Description: glycine-rich protein GRP22 - rape >gi|17821|emb|CAA78762| (Z15045) glycine-rich_protein_(aa1-291) [Brassica napus]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3957: from 95 to 108

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3958
- Ceres seq_id 1578854
- Location of start within SEQ ID NO 3956: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L7/L12 C-terminal domain

- Location within SEQ ID NO 3958: from 100 to 158 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1729
- gi No. 322647
- Description: glycine-rich protein GRP22 - rape >gi|17821|emb|CAA78762| (Z15045) glycine-rich_protein_(aa1-291) [Brassica napus]
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3958: from 72 to 85

Maximum Length Sequence:
related to:
Clone IDs:
296946
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3959
- Ceres seq_id 1578861
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3960
- Ceres seq_id 1578862
- Location of start within SEQ ID NO 3959: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1730
- gi No. 3822367
- Description: (AF081810) LdOrf-132 [Lymantria dispar nucleopolyhedrovirus]
- % Identity: 78.6
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3960: from 114 to 127

Maximum Length Sequence:
related to:
Clone IDs:
297256
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3961
- Ceres seq_id 1578909
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3962
- Ceres seq_id 1578910
- Location of start within SEQ ID NO 3961: at 138 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1731
- gi No. 82114
- Description: wound-induced protein - tomato (fragment) >gi|19320|emb|CAA42537| (X59882) wound induced protein [Lycopersicon esculentum]
- % Identity: 87.5
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 3962: from 9 to 32

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3963

- Ceres seq_id 1578911
- Location of start within SEQ ID NO 3961: at 165 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1732
- gi No. 82114
- Description: wound-induced protein - tomato (fragment) >gi|19320|emb|CAA42537| (X59882) wound induced protein [Lycopersicon esculentum]
- % Identity: 87.5
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 3963: from 1 to 23

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3964
- Ceres seq_id 1578912
- Location of start within SEQ ID NO 3961: at 174 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1733
- gi No. 82114
- Description: wound-induced protein - tomato (fragment) >gi|19320|emb|CAA42537| (X59882) wound induced protein [Lycopersicon esculentum]
- % Identity: 87.5
- Alignment Length: 24
- Location of Alignment in SEQ ID NO 3964: from 1 to 20

Maximum Length Sequence:
related to:
Clone IDs:
297892

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 3965
- Ceres seq_id 1578945

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3966
- Ceres seq_id 1578946
- Location of start within SEQ ID NO 3965: at 198 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- KOW motif
- Location within SEQ ID NO 3966: from 49 to 106 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1734
- gi No. 3747050
- Description: (AF093540) ribosomal protein L26 [Zea mays]
- % Identity: 96.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3966: from 1 to 119

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 3967
- Ceres seq_id 1578947
- Location of start within SEQ ID NO 3965: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- KOW motif
- Location within SEQ ID NO 3967: from 20 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1735
- gi No. 3747050
- Description: (AF093540) ribosomal protein L26 [Zea mays]
- % Identity: 96.6
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 3967: from 1 to 90

Maximum Length Sequence:
related to:
Clone IDs:
298044
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3968
- Ceres seq_id 1578951
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3969
- Ceres seq_id 1578952
- Location of start within SEQ ID NO 3968: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3969: from 111 to 161 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1736
- gi No. 2894120
- Description: (AJ003220) extensin-like protein [Solanum tuberosum]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3969: from 122 to 132

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3970
- Ceres seq_id 1578953
- Location of start within SEQ ID NO 3968: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 3970: from 88 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1737
- gi No. 2894120
- Description: (AJ003220) extensin-like protein [Solanum tuberosum]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 3970: from 99 to 109

Maximum Length Sequence:
related to:
Clone IDs:
299532

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3971
- Ceres seq_id 1578993

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3972
- Ceres seq_id 1578994
- Location of start within SEQ ID NO 3971: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 3972: from 123 to 172 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3973
- Ceres seq_id 1578995
- Location of start within SEQ ID NO 3971: at 201 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- DnaJ domain
- Location within SEQ ID NO 3973: from 57 to 106 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
299851

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3974
- Ceres seq_id 1579004

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3975
- Ceres seq_id 1579005
- Location of start within SEQ ID NO 3974: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rieske iron-sulfur protein
- Location within SEQ ID NO 3975: from 120 to 226 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1738
- gi No. 3885886
- Description: (AF093631) Rieske Fe-S precursor protein [Oryza sativa]
- % Identity: 85.6
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 3975: from 41 to 226

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3976
- Ceres seq_id 1579006
- Location of start within SEQ ID NO 3974: at 114 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rieske iron-sulfur protein

- Location within SEQ ID NO 3976: from 83 to 189 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1739
- gi No. 3885886
- Description: (AF093631) Rieske Fe-S precursor protein [Oryza sativa]
- % Identity: 85.6
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 3976: from 4 to 189

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3977
- Ceres seq_id 1579007
- Location of start within SEQ ID NO 3974: at 213 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Rieske iron-sulfur protein
- Location within SEQ ID NO 3977: from 50 to 156 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1740
- gi No. 3885886
- Description: (AF093631) Rieske Fe-S precursor protein [Oryza sativa]
- % Identity: 85.6
- Alignment Length: 189
- Location of Alignment in SEQ ID NO 3977: from 1 to 156

Maximum Length Sequence:
related to:
Clone IDs:
299932
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3978
- Ceres seq_id 1579012
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3979
- Ceres seq_id 1579013
- Location of start within SEQ ID NO 3978: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1741
- gi No. 1346809
- Description: PATHOGEN-RELATED PROTEIN >gi|499074|emb|CAA34641| (X16648) pathogenesis related protein [Hordeum vulgare]
- % Identity: 79.1
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 3979: from 42 to 211

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3980
- Ceres seq_id 1579014
- Location of start within SEQ ID NO 3978: at 125 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 1742
- gi No. 1346809
- Description: PATHOGEN-RELATED PROTEIN >gi|499074|emb|CAA34641|
(X16648) pathogenesis related protein [Hordeum vulgare]
- % Identity: 79.1
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 3980: from 1 to 170

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3981
- Ceres seq_id 1579015
- Location of start within SEQ ID NO 3978: at 320 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1743
- gi No. 1346809
- Description: PATHOGEN-RELATED PROTEIN >gi|499074|emb|CAA34641|
(X16648) pathogenesis related protein [Hordeum vulgare]
- % Identity: 79.1
- Alignment Length: 172
- Location of Alignment in SEQ ID NO 3981: from 1 to 105

Maximum Length Sequence:
related to:
Clone IDs:
300116
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3982
- Ceres seq_id 1579027
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3983
- Ceres seq_id 1579028
- Location of start within SEQ ID NO 3982: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1744
- gi No. 2257756
- Description: (U82815) nucleolar histone deacetylase HD2-p39 [Zea mays] >gi|3650466 (AF026917) histone deacetylase HD2-p39 [Zea mays]
- % Identity: 94.2
- Alignment Length: 155
- Location of Alignment in SEQ ID NO 3983: from 1 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3984
- Ceres seq_id 1579029
- Location of start within SEQ ID NO 3982: at 464 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1745
- gi No. 2257756
- Description: (U82815) nucleolar histone deacetylase HD2-p39 [Zea mays] >gi|3650466 (AF026917) histone deacetylase HD2-p39 [Zea mays]
- % Identity: 95.1
- Alignment Length: 142

- Location of Alignment in SEQ ID NO 3984: from 32 to 171

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3985
- Ceres seq_id 1579030
- Location of start within SEQ ID NO 3982: at 476 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1746
- gi No. 2257756
- Description: (U82815) nucleolar histone deacetylase HD2-p39 [Zea mays] >gi|3650466 (AF026917) histone deacetylase HD2-p39 [Zea mays]
- % Identity: 95.1
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 3985: from 28 to 167

Maximum Length Sequence:
related to:
Clone IDs:
300265
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3986
- Ceres seq_id 1579042
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3987
- Ceres seq_id 1579043
- Location of start within SEQ ID NO 3986: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1747
- gi No. 2134213
- Description: protamine I - American alligator
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3987: from 121 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3988
- Ceres seq_id 1579044
- Location of start within SEQ ID NO 3986: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Hsp20/alpha crystallin family
- Location within SEQ ID NO 3988: from 75 to 190 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3989
- Ceres seq_id 1579045
- Location of start within SEQ ID NO 3986: at 202 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1748

- gi No. 2134213
- Description: protamine I - American alligator
- % Identity: 71.4
- Alignment Length: 14
- Location of Alignment in SEQ ID NO 3989: from 54 to 67

Maximum Length Sequence:
related to:
Clone IDs:
300600
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3990
- Ceres seq_id 1579078

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3991
- Ceres seq_id 1579079
- Location of start within SEQ ID NO 3990: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3992
- Ceres seq_id 1579080
- Location of start within SEQ ID NO 3990: at 155 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3992: from 2 to 126 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1749
- gi No. 3914423
- Description: PROFILIN 4 >gi|2642324 (AF032370) profilin [Zea mays]
- % Identity: 95.4
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 3992: from 1 to 130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3993
- Ceres seq_id 1579081
- Location of start within SEQ ID NO 3990: at 302 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Profilins
- Location within SEQ ID NO 3993: from 1 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1750
- gi No. 3914423
- Description: PROFILIN 4 >gi|2642324 (AF032370) profilin [Zea mays]
- % Identity: 95.4
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 3993: from 1 to 81

Maximum Length Sequence:

related to:
Clone IDs:
300700
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3994
- Ceres seq_id 1579082
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3995
- Ceres seq_id 1579083
- Location of start within SEQ ID NO 3994: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3996
- Ceres seq_id 1579084
- Location of start within SEQ ID NO 3994: at 171 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Pentapeptide repeats (8 copies)
- Location within SEQ ID NO 3996: from 101 to 135 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
300949
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 3997
- Ceres seq_id 1579106
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3998
- Ceres seq_id 1579107
- Location of start within SEQ ID NO 3997: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 3998: from 20 to 213 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 3999
- Ceres seq_id 1579108
- Location of start within SEQ ID NO 3997: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glutathione S-transferases.
- Location within SEQ ID NO 3999: from 9 to 202 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:

301019
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4000
- Ceres seq_id 1579116
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4001
- Ceres seq_id 1579117
- Location of start within SEQ ID NO 4000: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1751
- gi No. 1350977
- Description: PROBABLE 40S RIBOSOMAL PROTEIN S30
- % Identity: 91.7
- Alignment Length: 60
- Location of Alignment in SEQ ID NO 4001: from 25 to 84

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4002
- Ceres seq_id 1579118
- Location of start within SEQ ID NO 4000: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4003
- Ceres seq_id 1579119
- Location of start within SEQ ID NO 4000: at 68 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1752
- gi No. 1350977
- Description: PROBABLE 40S RIBOSOMAL PROTEIN S30
- % Identity: 91.7
- Alignment Length: 60
- Location of Alignment in SEQ ID NO 4003: from 3 to 62

Maximum Length Sequence:
related to:
Clone IDs:
301126
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4004
- Ceres seq_id 1579128
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4005
- Ceres seq_id 1579129
- Location of start within SEQ ID NO 4004: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4006

- Ceres seq_id 1579130
- Location of start within SEQ ID NO 4004: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4007
- Ceres seq_id 1579131
- Location of start within SEQ ID NO 4004: at 100 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L31e
- Location within SEQ ID NO 4007: from 10 to 104 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1753
- gi No. 1173027
- Description: 60S RIBOSOMAL PROTEIN L31 >gi|915313 (U23784) ribosomal protein L31 [Nicotiana glutinosa]
- % Identity: 88.5
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 4007: from 10 to 122

Maximum Length Sequence:
related to:
Clone IDs:
301598
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4008
- Ceres seq_id 1579182

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4009
- Ceres seq_id 1579183
- Location of start within SEQ ID NO 4008: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 4009: from 28 to 88 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1754
- gi No. 733458
- Description: (U23190) chlorophyll a/b-binding apoprotein CP24 precursor [Zeamays]
- % Identity: 88.2
- Alignment Length: 87
- Location of Alignment in SEQ ID NO 4009: from 14 to 88

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4010
- Ceres seq_id 1579184
- Location of start within SEQ ID NO 4008: at 343 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 4010: from 1 to 148 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1755
- gi No. 115834
- Description: CHLOROPHYLL A-B BINDING PROTEIN CP24 10B PRECURSOR (CAB-10B) (LHCP) >gi|100202|pir||S11878 chlorophyll a/b-binding protein Cab10B - tomato >gi|170400 (M32606) chlorophyll b-binding protein [Lycopersicon esculentum]
- % Identity: 81.6
- Alignment Length: 174
- Location of Alignment in SEQ ID NO 4010: from 1 to 149

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4011
- Ceres seq_id 1579185
- Location of start within SEQ ID NO 4008: at 472 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Chlorophyll A-B binding proteins
- Location within SEQ ID NO 4011: from 1 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1756
- gi No. 115834
- Description: CHLOROPHYLL A-B BINDING PROTEIN CP24 10B PRECURSOR (CAB-10B) (LHCP) >gi|100202|pir||S11878 chlorophyll a/b-binding protein Cab10B - tomato >gi|170400 (M32606) chlorophyll b-binding protein [Lycopersicon esculentum]
- % Identity: 81.6
- Alignment Length: 174
- Location of Alignment in SEQ ID NO 4011: from 1 to 106

Maximum Length Sequence:
related to:
Clone IDs:
302115
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4012
- Ceres seq_id 1579249
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4013
- Ceres seq_id 1579250
- Location of start within SEQ ID NO 4012: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FKBP-type peptidyl-prolyl cis-trans isomerases
- Location within SEQ ID NO 4013: from 60 to 150 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4014
- Ceres seq_id 1579251
- Location of start within SEQ ID NO 4012: at 147 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FKBP-type peptidyl-prolyl cis-trans isomerases
- Location within SEQ ID NO 4014: from 12 to 102 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4015
- Ceres seq_id 1579252
- Location of start within SEQ ID NO 4012: at 150 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- FKBP-type peptidyl-prolyl cis-trans isomerases
- Location within SEQ ID NO 4015: from 11 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
302177
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4016
- Ceres seq_id 1579256

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4017
- Ceres seq_id 1579257
- Location of start within SEQ ID NO 4016: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4018
- Ceres seq_id 1579258
- Location of start within SEQ ID NO 4016: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1757
- gi No. 2493852
- Description: CYTOCHROME C OXIDASE POLYPEPTIDE VC >gi|1070356|emb|CAA92107| (Z68091) cytochrome c oxidase, Vc subunit [Hordeum vulgare]
- % Identity: 95.2
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 4018: from 28 to 90

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4019
- Ceres seq_id 1579259
- Location of start within SEQ ID NO 4016: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1758
- gi No. 2493852
- Description: CYTOCHROME C OXIDASE POLYPEPTIDE VC >gi|1070356|emb|CAA92107| (Z68091) cytochrome c oxidase, Vc subunit [Hordeum vulgare]

- % Identity: 95.2
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 4019: from 1 to 63

Maximum Length Sequence:
related to:
Clone IDs:
302267
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4020
- Ceres seq_id 1579263
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4021
- Ceres seq_id 1579264
- Location of start within SEQ ID NO 4020: at 115 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 4021: from 25 to 151 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4022
- Ceres seq_id 1579265
- Location of start within SEQ ID NO 4020: at 163 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 4022: from 9 to 135 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4023
- Ceres seq_id 1579266
- Location of start within SEQ ID NO 4020: at 256 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sugar (and other) transporter
- Location within SEQ ID NO 4023: from 1 to 104 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
302593
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4024
- Ceres seq_id 1579270
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4025
- Ceres seq_id 1579271
- Location of start within SEQ ID NO 4024: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- short chain dehydrogenase
- Location within SEQ ID NO 4025: from 34 to 155 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4026
- Ceres seq_id 1579273
- Location of start within SEQ ID NO 4024: at 28 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- short chain dehydrogenase
- Location within SEQ ID NO 4026: from 25 to 146 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
302875
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4027
- Ceres seq_id 1579274

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4028
- Ceres seq_id 1579275
- Location of start within SEQ ID NO 4027: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4029
- Ceres seq_id 1579276
- Location of start within SEQ ID NO 4027: at 94 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Photosystem I psaG / psaK
- Location within SEQ ID NO 4029: from 50 to 129 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1759
- gi No. 548605
- Description: PHOTOSYSTEM I REACTION CENTRE SUBUNIT X PRECURSOR (LIGHT-HARVESTING COMPLEX I 7 KD PROTEIN) (PSI-K) >gi|539055|pir||A48527 photosystem I protein psaK precursor - barley >gi|304220 (L12707) photosystem I PSI-K subunit [Hordeum vulgare]
- % Identity: 85.8
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 4029: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4030
- Ceres seq_id 1579277
- Location of start within SEQ ID NO 4027: at 259 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Photosystem I psaG / psaK
- Location within SEQ ID NO 4030: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1760
- gi No. 548605
- Description: PHOTOSYSTEM I REACTION CENTRE SUBUNIT X PRECURSOR (LIGHT-HARVESTING COMPLEX I 7 KD PROTEIN) (PSI-K) >gi|539055|pir||A48527 photosystem I protein psaK precursor - barley >gi|304220 (L12707) photosystem I PSI-K subunit [Hordeum vulgare]
- % Identity: 85.8
- Alignment Length: 136
- Location of Alignment in SEQ ID NO 4030: from 1 to 79

Maximum Length Sequence:
related to:
Clone IDs:
303071
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4031
- Ceres seq_id 1579278

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4032
- Ceres seq_id 1579279
- Location of start within SEQ ID NO 4031: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4033
- Ceres seq_id 1579280
- Location of start within SEQ ID NO 4031: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 4033: from 77 to 176 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4034
- Ceres seq_id 1579281
- Location of start within SEQ ID NO 4031: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 4034: from 50 to 149 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
303456
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4035
- Ceres seq_id 1579303

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4036
- Ceres seq_id 1579304
- Location of start within SEQ ID NO 4035: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic ribosomal protein L18
- Location within SEQ ID NO 4036: from 14 to 165 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1761
- gi No. 1172977
- Description: 60S RIBOSOMAL PROTEIN L18 >gi|606970 (U15741) cytoplasmic ribosomal protein L18 [Arabidopsis thaliana]
- % Identity: 80.9
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4036: from 14 to 165

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4037
- Ceres seq_id 1579305
- Location of start within SEQ ID NO 4035: at 285 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic ribosomal protein L18
- Location within SEQ ID NO 4037: from 1 to 134 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1762
- gi No. 1172977
- Description: 60S RIBOSOMAL PROTEIN L18 >gi|606970 (U15741) cytoplasmic ribosomal protein L18 [Arabidopsis thaliana]
- % Identity: 80.9
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4037: from 1 to 134

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4038
- Ceres seq_id 1579306
- Location of start within SEQ ID NO 4035: at 315 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic ribosomal protein L18
- Location within SEQ ID NO 4038: from 1 to 124 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1763
- gi No. 1172977
- Description: 60S RIBOSOMAL PROTEIN L18 >gi|606970 (U15741) cytoplasmic ribosomal protein L18 [Arabidopsis thaliana]
- % Identity: 80.9
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4038: from 1 to 124

Maximum Length Sequence:
related to:
Clone IDs:
303464

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4039
- Ceres seq_id 1579307

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4040
- Ceres seq_id 1579308
- Location of start within SEQ ID NO 4039: at 70 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1764
- gi No. 100219
- Description: glycine-rich protein (clone uK-4) - tomato >gi|1345534|emb|CAA39225| (X55696) glycine-rich protein [Lycopersicon esculentum]
- % Identity: 73.7
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 4040: from 7 to 25

Maximum Length Sequence:
related to:
Clone IDs:
303909

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4041
- Ceres seq_id 1579351

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4042
- Ceres seq_id 1579352
- Location of start within SEQ ID NO 4041: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S26e
- Location within SEQ ID NO 4042: from 40 to 162 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1765
- gi No. 1350969
- Description: 40S RIBOSOMAL PROTEIN S26 (S31) >gi|971284|dbj|BAA07208| (D38011) ribosomal protein S31 [Oryza sativa]
- % Identity: 85.4
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 4042: from 40 to 169

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4043
- Ceres seq_id 1579353
- Location of start within SEQ ID NO 4041: at 120 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S26e
- Location within SEQ ID NO 4043: from 1 to 123 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1766
- gi No. 1350969
- Description: 40S RIBOSOMAL PROTEIN S26 (S31) >gi|971284|dbj|BAA07208| (D38011) ribosomal protein S31 [Oryza sativa]

- % Identity: 85.4
- Alignment Length: 130
- Location of Alignment in SEQ ID NO 4043: from 1 to 130

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4044
- Ceres seq_id 1579354
- Location of start within SEQ ID NO 4041: at 274 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
303930
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4045
- Ceres seq_id 1579355

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4046
- Ceres seq_id 1579356
- Location of start within SEQ ID NO 4045: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Amino acid permease
- Location within SEQ ID NO 4046: from 2 to 58 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4047
- Ceres seq_id 1579357
- Location of start within SEQ ID NO 4045: at 73 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4048
- Ceres seq_id 1579358
- Location of start within SEQ ID NO 4045: at 211 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
303971
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4049
- Ceres seq_id 1579359

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4050
- Ceres seq_id 1579360
- Location of start within SEQ ID NO 4049: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1767
- gi No. 121689
- Description: GAST1 PROTEIN PRECURSOR >gi|100217|pir||S22151 GAST1 protein - tomato >gi|19247|emb|CAA44807| (X63093) gast1 [Lycopersicon esculentum]
- % Identity: 70
- Alignment Length: 70
- Location of Alignment in SEQ ID NO 4050: from 93 to 162

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4051
- Ceres seq_id 1579361
- Location of start within SEQ ID NO 4049: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4052
- Ceres seq_id 1579362
- Location of start within SEQ ID NO 4049: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
304049
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4053
- Ceres seq_id 1579366
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4054
- Ceres seq_id 1579367
- Location of start within SEQ ID NO 4053: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1768
- gi No. 642484
- Description: (U16371) androgen receptor [Homo sapiens]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4054: from 121 to 132

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4055
- Ceres seq_id 1579368
- Location of start within SEQ ID NO 4053: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 1769
- gi No. 642484
- Description: (U16371) androgen receptor [Homo sapiens]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4055: from 92 to 103

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4056
- Ceres seq_id 1579369
- Location of start within SEQ ID NO 4053: at 186 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1770
- gi No. 642484
- Description: (U16371) androgen receptor [Homo sapiens]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4056: from 60 to 71

Maximum Length Sequence:
related to:
Clone IDs:
304237
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4057
- Ceres seq_id 1579382
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4058
- Ceres seq_id 1579383
- Location of start within SEQ ID NO 4057: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4059
- Ceres seq_id 1579384
- Location of start within SEQ ID NO 4057: at 60 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1771
- gi No. 3914465
- Description: PHOTOSYSTEM I REACTION CENTRE SUBUNIT VI PRECURSOR (LIGHT-HARVESTING COMPLEX I 11 KD PROTEIN) (PSI-H) >gi|2981207 (AF052076) photosystem I complex PsaH subunit precursor [Zea mays]
- % Identity: 100
- Alignment Length: 142
- Location of Alignment in SEQ ID NO 4059: from 1 to 142

Maximum Length Sequence:
related to:
Clone IDs:
304280
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4060

- Ceres seq_id 1579398

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4061
- Ceres seq_id 1579399
- Location of start within SEQ ID NO 4060: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Core histone H2A/H2B/H3/H4
- Location within SEQ ID NO 4061: from 29 to 153 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1772
- gi No. 729671
- Description: HISTONE H2A >gi|473603 (U08225) histone H2A [Zea mays]
- % Identity: 93.4
- Alignment Length: 151
- Location of Alignment in SEQ ID NO 4061: from 29 to 179

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4062
- Ceres seq_id 1579400
- Location of start within SEQ ID NO 4060: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4063
- Ceres seq_id 1579401
- Location of start within SEQ ID NO 4060: at 69 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
304287
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4064
- Ceres seq_id 1579402

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4065
- Ceres seq_id 1579403
- Location of start within SEQ ID NO 4064: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S24e
- Location within SEQ ID NO 4065: from 85 to 169 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1773
- gi No. 445612
- Description: ribosomal protein S19 [Solanum tuberosum]
- % Identity: 81.8
- Alignment Length: 132

- Location of Alignment in SEQ ID NO 4065: from 63 to 194

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4066
- Ceres seq_id 1579404
- Location of start within SEQ ID NO 4064: at 172 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S24e
- Location within SEQ ID NO 4066: from 28 to 112 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1774
- gi No. 445612
- Description: ribosomal protein S19 [Solanum tuberosum]
- % Identity: 81.8
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 4066: from 6 to 137

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4067
- Ceres seq_id 1579405
- Location of start within SEQ ID NO 4064: at 223 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S24e
- Location within SEQ ID NO 4067: from 11 to 95 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1775
- gi No. 445612
- Description: ribosomal protein S19 [Solanum tuberosum]
- % Identity: 81.8
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 4067: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
304376
243869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4068
- Ceres seq_id 1579406
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4069
- Ceres seq_id 1579407
- Location of start within SEQ ID NO 4068: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonuclease T2 family
- Location within SEQ ID NO 4069: from 85 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1776
- gi No. 1698670
- Description: (U66241) S-like RNase [Zea mays]
- % Identity: 97.7

- Alignment Length: 129
- Location of Alignment in SEQ ID NO 4069: from 7 to 131

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4070
- Ceres seq_id 1579408
- Location of start within SEQ ID NO 4068: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonuclease T2 family
- Location within SEQ ID NO 4070: from 79 to 124 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1777
- gi No. 1698670
- Description: (U66241) S-like RNase [Zea mays]
- % Identity: 97.7
- Alignment Length: 129
- Location of Alignment in SEQ ID NO 4070: from 1 to 125

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4071
- Ceres seq_id 1579409
- Location of start within SEQ ID NO 4068: at 36 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonuclease T2 family
- Location within SEQ ID NO 4071: from 74 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1778
- gi No. 1698670
- Description: (U66241) S-like RNase [Zea mays]
- % Identity: 97.7
- Alignment Length: 129
- Location of Alignment in SEQ ID NO 4071: from 1 to 120

Maximum Length Sequence:
related to:
Clone IDs:
305329
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4072
- Ceres seq_id 1579440

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4073
- Ceres seq_id 1579441
- Location of start within SEQ ID NO 4072: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4073: from 60 to 119 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1779
- gi No. 100210
- Description: extensin precursor (clone Tom L-4) - tomato
>gi|170444 (M76671) extensin (class II) [Lycopersicon esculentum]

- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4073: from 13 to 24

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4074
- Ceres seq_id 1579442
- Location of start within SEQ ID NO 4072: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4075
- Ceres seq_id 1579443
- Location of start within SEQ ID NO 4072: at 121 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
305464
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4076
- Ceres seq_id 1579447
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4077
- Ceres seq_id 1579448
- Location of start within SEQ ID NO 4076: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- dUTPase
- Location within SEQ ID NO 4077: from 65 to 194 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1780
- gi No. 5541665
- Description: (AL096859) dUTP pyrophosphatase-like protein [Arabidopsis thaliana]
- % Identity: 81.9
- Alignment Length: 144
- Location of Alignment in SEQ ID NO 4077: from 52 to 195

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4078
- Ceres seq_id 1579449
- Location of start within SEQ ID NO 4076: at 262 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- dUTPase
- Location within SEQ ID NO 4078: from 1 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1781
- gi No. 5541665

- Description: (AL096859) dUTP pyrophosphatase-like protein [Arabidopsis thaliana]
- % Identity: 81.9
- Alignment Length: 144
- Location of Alignment in SEQ ID NO 4078: from 1 to 108

Maximum Length Sequence:
related to:
Clone IDs:
305482
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4079
- Ceres seq_id 1579454
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4080
- Ceres seq_id 1579455
- Location of start within SEQ ID NO 4079: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1782
- gi No. 1352054
- Description: ATP SYNTHASE 6 KD SUBUNIT, MITOCHONDRIAL
- % Identity: 76
- Alignment Length: 25
- Location of Alignment in SEQ ID NO 4080: from 28 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4081
- Ceres seq_id 1579456
- Location of start within SEQ ID NO 4079: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
305485
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4082
- Ceres seq_id 1579457
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4083
- Ceres seq_id 1579458
- Location of start within SEQ ID NO 4082: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4084
- Ceres seq_id 1579459
- Location of start within SEQ ID NO 4082: at 261 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

- Alignment No. 1783
- gi No. 2425064
- Description: (AF019146) cysteine proteinase Mir2 [Zea mays]
- % Identity: 90.9
- Alignment Length: 33
- Location of Alignment in SEQ ID NO 4084: from 40 to 72

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4085
- Ceres seq_id 1579460
- Location of start within SEQ ID NO 4082: at 532 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
305518
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4086
- Ceres seq_id 1579468
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4087
- Ceres seq_id 1579469
- Location of start within SEQ ID NO 4086: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4088
- Ceres seq_id 1579470
- Location of start within SEQ ID NO 4086: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 4088: from 1 to 61 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1784
- gi No. 266578
- Description: METALLOTHIONEIN-LIKE PROTEIN 1 >gi|100898|pir||S17560 metallothionein-like protein - maize >gi|236730|bbs|57629 (S57628) metallothionein homologue [Zea mays, Peptide, 76 aa] [Zea mays] >gi|559536|emb|CAA57676| (X82186)
- % Identity: 96.7
- Alignment Length: 61
- Location of Alignment in SEQ ID NO 4088: from 1 to 61

Maximum Length Sequence:
related to:
Clone IDs:
305552
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4089
- Ceres seq_id 1579471
(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4090
- Ceres seq_id 1579472
- Location of start within SEQ ID NO 4089: at 78 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S19e
- Location within SEQ ID NO 4090: from 7 to 143 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1785
- gi No. 730456
- Description: 40S RIBOSOMAL PROTEIN S19
- % Identity: 95.8
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 4090: from 1 to 143

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4091
- Ceres seq_id 1579473
- Location of start within SEQ ID NO 4089: at 168 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S19e
- Location within SEQ ID NO 4091: from 1 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1786
- gi No. 730456
- Description: 40S RIBOSOMAL PROTEIN S19
- % Identity: 95.8
- Alignment Length: 143
- Location of Alignment in SEQ ID NO 4091: from 1 to 113

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4092
- Ceres seq_id 1579474
- Location of start within SEQ ID NO 4089: at 322 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
305992
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4093
- Ceres seq_id 1579490

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4094
- Ceres seq_id 1579491
- Location of start within SEQ ID NO 4093: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4095
- Ceres seq_id 1579492
- Location of start within SEQ ID NO 4093: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1787
- gi No. 544867
- Description: beta-glucosidase [Hordeum vulgare=barley, Sofia, Peptide Partial, 41 aa, segment 1 of 6]
- % Identity: 82.4
- Alignment Length: 17
- Location of Alignment in SEQ ID NO 4095: from 33 to 49

Maximum Length Sequence:
related to:
Clone IDs:
306077
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4096
- Ceres seq_id 1579493
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4097
- Ceres seq_id 1579494
- Location of start within SEQ ID NO 4096: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1788
- gi No. 1841355
- Description: (D85381) cytochrome c oxidase subunit Vb precursor [Oryza sativa]
- % Identity: 71.7
- Alignment Length: 53
- Location of Alignment in SEQ ID NO 4097: from 23 to 71

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4098
- Ceres seq_id 1579495
- Location of start within SEQ ID NO 4096: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1789
- gi No. 465445
- Description: PROBABLE NUCLEAR ANTIGEN >gi|418708|pir||B45344 probable nuclear antigen - suid herpesvirus 1 (strain Kaplan) >gi|334072 (M34651) ORF-3 protein [Pseudorabies virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4098: from 20 to 30

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4099
- Ceres seq_id 1579496
- Location of start within SEQ ID NO 4096: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1790
- gi No. 1841355
- Description: (D85381) cytochrome c oxidase subunit Vb precursor [Oryza sativa]
- % Identity: 71.7
- Alignment Length: 53
- Location of Alignment in SEQ ID NO 4099: from 1 to 49

Maximum Length Sequence:
related to:
Clone IDs:
306390

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4100
- Ceres seq_id 1579503

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4101
- Ceres seq_id 1579504
- Location of start within SEQ ID NO 4100: at 75 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 4101: from 68 to 125 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4102
- Ceres seq_id 1579505
- Location of start within SEQ ID NO 4100: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 4102: from 66 to 123 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4103
- Ceres seq_id 1579506
- Location of start within SEQ ID NO 4100: at 84 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal L29 protein
- Location within SEQ ID NO 4103: from 65 to 122 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
339910

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4104
- Ceres seq_id 1579543

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4105
- Ceres seq_id 1579544
- Location of start within SEQ ID NO 4104: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L23
- Location within SEQ ID NO 4105: from 92 to 151 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1791
- gi No. 585876
- Description: 60S RIBOSOMAL PROTEIN L23A (L25) >gi|1084424|pir||S48026 ribosomal protein L25 - common tobacco >gi|310935 (L18908) 60S ribosomal protein L25 [Nicotiana tabacum]
- % Identity: 82.6
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 4105: from 23 to 151

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4106
- Ceres seq_id 1579545
- Location of start within SEQ ID NO 4104: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein L23
- Location within SEQ ID NO 4106: from 70 to 129 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1792
- gi No. 585876
- Description: 60S RIBOSOMAL PROTEIN L23A (L25) >gi|1084424|pir||S48026 ribosomal protein L25 - common tobacco >gi|310935 (L18908) 60S ribosomal protein L25 [Nicotiana tabacum]
- % Identity: 82.6
- Alignment Length: 132
- Location of Alignment in SEQ ID NO 4106: from 1 to 129

Maximum Length Sequence:
related to:
Clone IDs:
339920
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4107
- Ceres seq_id 1579546

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4108
- Ceres seq_id 1579547
- Location of start within SEQ ID NO 4107: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1793
- gi No. 1173218
- Description: 40S RIBOSOMAL PROTEIN S15A >gi|440824 (L27461) ribosomal protein S15 [Arabidopsis thaliana] >gi|2150130 (AF001412) cytoplasmic ribosomal protein S15a [Arabidopsis thaliana]
- % Identity: 93.3

- Alignment Length: 45
- Location of Alignment in SEQ ID NO 4108: from 29 to 73

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4109
- Ceres seq_id 1579548
- Location of start within SEQ ID NO 4107: at 87 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1794
- gi No. 1173218
- Description: 40S RIBOSOMAL PROTEIN S15A >gi|440824 (L27461) ribosomal protein S15 [Arabidopsis thaliana] >gi|2150130 (AF001412) cytoplasmic ribosomal protein S15a [Arabidopsis thaliana]
- % Identity: 93.3
- Alignment Length: 45
- Location of Alignment in SEQ ID NO 4109: from 1 to 45

Maximum Length Sequence:
related to:
Clone IDs:
340020
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4110
- Ceres seq_id 1579569
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4111
- Ceres seq_id 1579570
- Location of start within SEQ ID NO 4110: at 118 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4112
- Ceres seq_id 1579571
- Location of start within SEQ ID NO 4110: at 203 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Mucin-like glycoprotein
- Location within SEQ ID NO 4112: from 1 to 75 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340039
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4113
- Ceres seq_id 1579576
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4114
- Ceres seq_id 1579577
- Location of start within SEQ ID NO 4113: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 4114: from 27 to 130 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4115
- Ceres seq_id 1579578
- Location of start within SEQ ID NO 4113: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4116
- Ceres seq_id 1579579
- Location of start within SEQ ID NO 4113: at 79 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 60s Acidic ribosomal protein
- Location within SEQ ID NO 4116: from 1 to 104 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340055
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4117
- Ceres seq_id 1579584
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4118
- Ceres seq_id 1579585
- Location of start within SEQ ID NO 4117: at 105 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 4118: from 8 to 72 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1795
- gi No. 134126
- Description: SMALL NUCLEAR RIBONUCLEOPROTEIN E HOMOLOG C29 >gi|19573|emb|CAA44975| (X63376) snRNP-related protein [Medicago sativa]
- % Identity: 82.9
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 4118: from 1 to 76

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4119
- Ceres seq_id 1579586
- Location of start within SEQ ID NO 4117: at 144 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Sm protein
- Location within SEQ ID NO 4119: from 1 to 59 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1796
- gi No. 134126
- Description: SMALL NUCLEAR RIBONUCLEOPROTEIN E HOMOLOG C29 >gi|19573|emb|CAA44975| (X63376) snRNP-related protein [Medicago sativa]
- % Identity: 82.9
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 4119: from 1 to 63

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4120
- Ceres seq_id 1579587
- Location of start within SEQ ID NO 4117: at 219 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1797
- gi No. 134126
- Description: SMALL NUCLEAR RIBONUCLEOPROTEIN E HOMOLOG C29 >gi|19573|emb|CAA44975| (X63376) snRNP-related protein [Medicago sativa]
- % Identity: 82.9
- Alignment Length: 76
- Location of Alignment in SEQ ID NO 4120: from 1 to 38

Maximum Length Sequence:
related to:
Clone IDs:
340102
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4121
- Ceres seq_id 1579596
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4122
- Ceres seq_id 1579597
- Location of start within SEQ ID NO 4121: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1798
- gi No. 3335148
- Description: (AF055376) short form transcription factor C-MAF [Homo sapiens]
- % Identity: 73.3
- Alignment Length: 15
- Location of Alignment in SEQ ID NO 4122: from 146 to 160

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4123
- Ceres seq_id 1579598
- Location of start within SEQ ID NO 4121: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1799
- gi No. 134780

- Description: SPORE COAT PROTEIN SP96 >gi|84145|pir||S07638 spore coat protein SP96 precursor - slime mold (Dictyostelium discoideum) >gi|295736|emb|CAA34508| (X16491) spore coat protein sp96 [Dictyostelium discoideum]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 4123: from 33 to 53

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4124
- Ceres seq_id 1579599
- Location of start within SEQ ID NO 4121: at 48 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1800
- gi No. 134780
- Description: SPORE COAT PROTEIN SP96 >gi|84145|pir||S07638 spore coat protein SP96 precursor - slime mold (Dictyostelium discoideum) >gi|295736|emb|CAA34508| (X16491) spore coat protein sp96 [Dictyostelium discoideum]
- % Identity: 71.4
- Alignment Length: 21
- Location of Alignment in SEQ ID NO 4124: from 18 to 38

Maximum Length Sequence:
related to:
Clone IDs:
340151
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4125
- Ceres seq_id 1579610

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4126
- Ceres seq_id 1579611
- Location of start within SEQ ID NO 4125: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4127
- Ceres seq_id 1579612
- Location of start within SEQ ID NO 4125: at 365 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1801
- gi No. 2231312
- Description: (U75603) AtRab18 [Arabidopsis thaliana]
- % Identity: 74.3
- Alignment Length: 113
- Location of Alignment in SEQ ID NO 4127: from 1 to 37

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4128
- Ceres seq_id 1579613
- Location of start within SEQ ID NO 4125: at 393 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340154
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4129
- Ceres seq_id 1579614
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4130
- Ceres seq_id 1579615
- Location of start within SEQ ID NO 4129: at 108 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 4130: from 41 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1802
- gi No. 3417455
- Description: (AB016810) ferredoxin [Zea mays]
- % Identity: 96.7
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 4130: from 1 to 118

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4131
- Ceres seq_id 1579616
- Location of start within SEQ ID NO 4129: at 192 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 2Fe-2S iron-sulfur cluster binding domains
- Location within SEQ ID NO 4131: from 13 to 90 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1803
- gi No. 3417455
- Description: (AB016810) ferredoxin [Zea mays]
- % Identity: 96.7
- Alignment Length: 120
- Location of Alignment in SEQ ID NO 4131: from 1 to 90

Maximum Length Sequence:
related to:
Clone IDs:
340166
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4132
- Ceres seq_id 1579618
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4133
- Ceres seq_id 1579619
- Location of start within SEQ ID NO 4132: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonucleotide reductases
- Location within SEQ ID NO 4133: from 14 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1804
- gi No. 1710401
- Description: RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (RIBONUCLEOTIDE REDUCTASE) (R2 SUBUNIT) >gi|1044912|emb|CAA63194| (X92443) ribonucleotide reductase R2 [Nicotiana tabacum]
- % Identity: 86.4
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 4133: from 14 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4134
- Ceres seq_id 1579620
- Location of start within SEQ ID NO 4132: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4135
- Ceres seq_id 1579621
- Location of start within SEQ ID NO 4132: at 95 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribonucleotide reductases
- Location within SEQ ID NO 4135: from 1 to 107 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1805
- gi No. 1710401
- Description: RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (RIBONUCLEOTIDE REDUCTASE) (R2 SUBUNIT) >gi|1044912|emb|CAA63194| (X92443) ribonucleotide reductase R2 [Nicotiana tabacum]
- % Identity: 86.4
- Alignment Length: 125
- Location of Alignment in SEQ ID NO 4135: from 1 to 107

Maximum Length Sequence:
related to:
Clone IDs:
340198
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4136
- Ceres seq_id 1579629

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4137
- Ceres seq_id 1579630
- Location of start within SEQ ID NO 4136: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 4137: from 44 to 113 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1806
- gi No. 5262856
- Description: (AJ238095) Lsm3 protein [Homo sapiens]
- % Identity: 85.7
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 4137: from 39 to 101

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4138
- Ceres seq_id 1579631
- Location of start within SEQ ID NO 4136: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sm protein
- Location within SEQ ID NO 4138: from 15 to 84 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1807
- gi No. 5262856
- Description: (AJ238095) Lsm3 protein [Homo sapiens]
- % Identity: 85.7
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 4138: from 10 to 72

Maximum Length Sequence:
related to:
Clone IDs:
340212
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4139
- Ceres seq_id 1579632

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4140
- Ceres seq_id 1579633
- Location of start within SEQ ID NO 4139: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4141
- Ceres seq_id 1579634
- Location of start within SEQ ID NO 4139: at 123 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S4/S9
- Location within SEQ ID NO 4141: from 9 to 94 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1808
- gi No. 3088337
- Description: (AB007150) ribosomal protein S9 [Homo sapiens]
- % Identity: 71.4
- Alignment Length: 63
- Location of Alignment in SEQ ID NO 4141: from 12 to 74

Maximum Length Sequence:

related to:
Clone IDs:
340221
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4142
- Ceres seq_id 1579635
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4143
- Ceres seq_id 1579636
- Location of start within SEQ ID NO 4142: at 91 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1809
- gi No. 2058524
- Description: (U78522) histidine secretory acid phosphatase [Leishmania donovani]
- % Identity: 75
- Alignment Length: 16
- Location of Alignment in SEQ ID NO 4143: from 15 to 30

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4144
- Ceres seq_id 1579637
- Location of start within SEQ ID NO 4142: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4145
- Ceres seq_id 1579638
- Location of start within SEQ ID NO 4142: at 190 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340261
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4146
- Ceres seq_id 1579643
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4147
- Ceres seq_id 1579644
- Location of start within SEQ ID NO 4146: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1810
- gi No. 4758388
- Description: ref|NP_004465.1|pFKHL17| forkhead (Drosophila)-like 17 >gi|3090887 (AF042832) forkhead-related transcription factor FREAC-9 [Homo sapiens]
- % Identity: 76.9

- Alignment Length: 13
- Location of Alignment in SEQ ID NO 4147: from 32 to 44

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4148
- Ceres seq_id 1579645
- Location of start within SEQ ID NO 4146: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1811
- gi No. 3025832
- Description: (AF055985) pyrrolidone-rich antigen [Onchocerca volvulus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4148: from 112 to 122

Maximum Length Sequence:
related to:
Clone IDs:
340279
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4149
- Ceres seq_id 1579646
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4150
- Ceres seq_id 1579647
- Location of start within SEQ ID NO 4149: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 4150: from 1 to 121 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4151
- Ceres seq_id 1579648
- Location of start within SEQ ID NO 4149: at 119 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Protein phosphatase 2C
- Location within SEQ ID NO 4151: from 1 to 82 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340442
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4152
- Ceres seq_id 1579700
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4153
- Ceres seq_id 1579701
- Location of start within SEQ ID NO 4152: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4154
- Ceres seq_id 1579702
- Location of start within SEQ ID NO 4152: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4154: from 8 to 71 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1812
- gi No. 2702442
- Description: (AF038614) contains similarity to Lentinula edodes MFBA=234.5 kda mature fruiting body adhesion protein (GB:S75826) [Caenorhabditis elegans]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4154: from 7 to 18

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4155
- Ceres seq_id 1579703
- Location of start within SEQ ID NO 4152: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340477
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4156
- Ceres seq_id 1579709
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4157
- Ceres seq_id 1579710
- Location of start within SEQ ID NO 4156: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1813
- gi No. 1429226
- Description: (X98861) TFIIA [Arabidopsis thaliana]
- % Identity: 77.8
- Alignment Length: 73
- Location of Alignment in SEQ ID NO 4157: from 9 to 80

Maximum Length Sequence:
related to:
Clone IDs:
340497
(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 4158
- Ceres seq_id 1579711

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4159
- Ceres seq_id 1579712
- Location of start within SEQ ID NO 4158: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4159: from 58 to 100 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4160
- Ceres seq_id 1579713
- Location of start within SEQ ID NO 4158: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4161
- Ceres seq_id 1579714
- Location of start within SEQ ID NO 4158: at 46 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4161: from 43 to 85 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340599

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4162
- Ceres seq_id 1579724

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4163
- Ceres seq_id 1579725
- Location of start within SEQ ID NO 4162: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 4163: from 14 to 42 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1814
- gi No. 3309086
- Description: (AF076253) calcineurin B-like protein 3 [Arabidopsis thaliana] >gi|4938495|emb|CAB43853.1| (AL078465) calcineurin B-like protein 3 [Arabidopsis thaliana]
- % Identity: 85.8
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 4163: from 1 to 106

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4164
- Ceres seq_id 1579726
- Location of start within SEQ ID NO 4162: at 38 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- EF hand
- Location within SEQ ID NO 4164: from 2 to 30 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1815
- gi No. 3309086
- Description: (AF076253) calcineurin B-like protein 3 [Arabidopsis thaliana] >gi|4938495|emb|CAB43853.1| (AL078465) calcineurin B-like protein 3 [Arabidopsis thaliana]
- % Identity: 85.8
- Alignment Length: 106
- Location of Alignment in SEQ ID NO 4164: from 1 to 94

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4165
- Ceres seq_id 1579727
- Location of start within SEQ ID NO 4162: at 81 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340758
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4166
- Ceres seq_id 1579756
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4167
- Ceres seq_id 1579757
- Location of start within SEQ ID NO 4166: at 116 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Metallothionein
- Location within SEQ ID NO 4167: from 1 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1816
- gi No. 2497903
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1752831|dbj|BAA14038.1| (D89931) metallothionein-like protein [Oryza sativa] >gi|1815628 (U43530) metallothionein-like type 2 [Oryza sativa]
- % Identity: 80
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 4167: from 1 to 74

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4168
- Ceres seq_id 1579758
- Location of start within SEQ ID NO 4166: at 188 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1817
- gi No. 2497903
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1752831|dbj|BAA14038.1| (D89931) metallothionein-like protein [Oryza sativa] >gi|1815628 (U43530) metallothionein-like type 2 [Oryza sativa]
- % Identity: 80
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 4168: from 1 to 50

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4169
- Ceres seq_id 1579759
- Location of start within SEQ ID NO 4166: at 200 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1818
- gi No. 2497903
- Description: METALLOTHIONEIN-LIKE PROTEIN TYPE 2 >gi|1752831|dbj|BAA14038.1| (D89931) metallothionein-like protein [Oryza sativa] >gi|1815628 (U43530) metallothionein-like type 2 [Oryza sativa]
- % Identity: 80
- Alignment Length: 81
- Location of Alignment in SEQ ID NO 4169: from 1 to 46

Maximum Length Sequence:
related to:
Clone IDs:
340834

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 4170
- Ceres seq_id 1579776

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4171
- Ceres seq_id 1579777
- Location of start within SEQ ID NO 4170: at 67 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Thaumatin family
- Location within SEQ ID NO 4171: from 33 to 122 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4172
- Ceres seq_id 1579778
- Location of start within SEQ ID NO 4170: at 110 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4173
- Ceres seq_id 1579779

- Location of start within SEQ ID NO 4170: at 130 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 4173: from 12 to 101 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
340910
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4174
- Ceres seq_id 1579794
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4175
- Ceres seq_id 1579795
- Location of start within SEQ ID NO 4174: at 170 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4176
- Ceres seq_id 1579796
- Location of start within SEQ ID NO 4174: at 178 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4176: from 23 to 77 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1819
- gi No. 544370
- Description: GAR1 PROTEIN >gi|422064|pir||S33691 GAR1 protein - fission yeast (Schizosaccharomyces pombe) >gi|297009|emb|CAA79628| (Z19576) snoRNP protein GAR1 [Schizosaccharomyces pombe] >gi|1799516|dbj|BAA19143| (AB000537) [Schizosaccharomyces pombe]
- % Identity: 78.9
- Alignment Length: 19
- Location of Alignment in SEQ ID NO 4176: from 22 to 39

Maximum Length Sequence:
related to:
Clone IDs:
340917
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4177
- Ceres seq_id 1579801
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4178
- Ceres seq_id 1579802
- Location of start within SEQ ID NO 4177: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S3, C-terminal domain.

- Location within SEQ ID NO 4178: from 98 to 155 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1820
- gi No. 1173253
- Description: 40S RIBOSOMAL PROTEIN S3 >gi|543317|pir||S41170 ribosomal protein S3 - mouse >gi|57728|emb|CAA35916| (X51536) ribosomal protein S3 (AA 1-243) [Rattus rattus] >gi|439522|emb|CAA54167| (X76772) ribosomal protein S3 [Mus musculus]
- % Identity: 88.8
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4178: from 4 to 155

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4179
- Ceres seq_id 1579803
- Location of start within SEQ ID NO 4177: at 53 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S3, C-terminal domain.
- Location within SEQ ID NO 4179: from 81 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1821
- gi No. 1173253
- Description: 40S RIBOSOMAL PROTEIN S3 >gi|543317|pir||S41170 ribosomal protein S3 - mouse >gi|57728|emb|CAA35916| (X51536) ribosomal protein S3 (AA 1-243) [Rattus rattus] >gi|439522|emb|CAA54167| (X76772) ribosomal protein S3 [Mus musculus]
- % Identity: 88.8
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4179: from 1 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4180
- Ceres seq_id 1579804
- Location of start within SEQ ID NO 4177: at 113 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Ribosomal protein S3, C-terminal domain.
- Location within SEQ ID NO 4180: from 61 to 118 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1822
- gi No. 1173253
- Description: 40S RIBOSOMAL PROTEIN S3 >gi|543317|pir||S41170 ribosomal protein S3 - mouse >gi|57728|emb|CAA35916| (X51536) ribosomal protein S3 (AA 1-243) [Rattus rattus] >gi|439522|emb|CAA54167| (X76772) ribosomal protein S3 [Mus musculus]
- % Identity: 88.8
- Alignment Length: 152
- Location of Alignment in SEQ ID NO 4180: from 1 to 118

Maximum Length Sequence:
related to:
Clone IDs:
340976
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4181

- Ceres seq_id 1579820

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4182
- Ceres seq_id 1579821
- Location of start within SEQ ID NO 4181: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Sperm histone P2
- Location within SEQ ID NO 4182: from 37 to 114 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1823
- gi No. 2134206
- Description: protamine I-1 - painted turtle
- % Identity: 73.9
- Alignment Length: 23
- Location of Alignment in SEQ ID NO 4182: from 70 to 91

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4183
- Ceres seq_id 1579822
- Location of start within SEQ ID NO 4181: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1824
- gi No. 1628459
- Description: (Y08774) Men-2 [Silene latifolia]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4183: from 88 to 99

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4184
- Ceres seq_id 1579823
- Location of start within SEQ ID NO 4181: at 207 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1825
- gi No. 2134206
- Description: protamine I-1 - painted turtle
- % Identity: 73.9
- Alignment Length: 23
- Location of Alignment in SEQ ID NO 4184: from 2 to 23

Maximum Length Sequence:
related to:
Clone IDs:
340990

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4185
- Ceres seq_id 1579828

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4186
- Ceres seq_id 1579829
- Location of start within SEQ ID NO 4185: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4187
- Ceres seq_id 1579830
- Location of start within SEQ ID NO 4185: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1826
- gi No. 124178
- Description: IMMEDIATE-EARLY PROTEIN IE180 >gi|59968|emb|CAA33214| (X15120) immediate-early protein (AA 1-1460) [Pseudorabies virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4187: from 133 to 143

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4188
- Ceres seq_id 1579831
- Location of start within SEQ ID NO 4185: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1827
- gi No. 124178
- Description: IMMEDIATE-EARLY PROTEIN IE180 >gi|59968|emb|CAA33214| (X15120) immediate-early protein (AA 1-1460) [Pseudorabies virus]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4188: from 92 to 102

Maximum Length Sequence:
related to:
Clone IDs:
341017
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4189
- Ceres seq_id 1579836

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4190
- Ceres seq_id 1579837
- Location of start within SEQ ID NO 4189: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 4190: from 72 to 122 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4191
- Ceres seq_id 1579838
- Location of start within SEQ ID NO 4189: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 4191: from 37 to 87 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4192
- Ceres seq_id 1579839
- Location of start within SEQ ID NO 4189: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Thaumatin family
- Location within SEQ ID NO 4192: from 16 to 66 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341154
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4193
- Ceres seq_id 1579865
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4194
- Ceres seq_id 1579866
- Location of start within SEQ ID NO 4193: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SIS domain
- Location within SEQ ID NO 4194: from 52 to 133 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4195
- Ceres seq_id 1579867
- Location of start within SEQ ID NO 4193: at 15 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SIS domain
- Location within SEQ ID NO 4195: from 48 to 129 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4196
- Ceres seq_id 1579868
- Location of start within SEQ ID NO 4193: at 21 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- SIS domain
- Location within SEQ ID NO 4196: from 46 to 127 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341198
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4197
- Ceres seq_id 1579873
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4198
- Ceres seq_id 1579874
- Location of start within SEQ ID NO 4197: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 4198: from 22 to 90 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4199
- Ceres seq_id 1579875
- Location of start within SEQ ID NO 4197: at 76 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- HMG (high mobility group) box
- Location within SEQ ID NO 4199: from 1 to 65 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4200
- Ceres seq_id 1579876
- Location of start within SEQ ID NO 4197: at 245 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341224
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4201
- Ceres seq_id 1579877
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4202
- Ceres seq_id 1579878
- Location of start within SEQ ID NO 4201: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4203
- Ceres seq_id 1579879

- Location of start within SEQ ID NO 4201: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4203: from 12 to 74 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341310
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4204
- Ceres seq_id 1579904
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4205
- Ceres seq_id 1579905
- Location of start within SEQ ID NO 4204: at 133 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidase family M41
- Location within SEQ ID NO 4205: from 46 to 113 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4206
- Ceres seq_id 1579906
- Location of start within SEQ ID NO 4204: at 169 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidase family M41
- Location within SEQ ID NO 4206: from 34 to 101 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4207
- Ceres seq_id 1579907
- Location of start within SEQ ID NO 4204: at 250 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidase family M41
- Location within SEQ ID NO 4207: from 7 to 74 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341347
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4208
- Ceres seq_id 1579911
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4209

- Ceres seq_id 1579912
- Location of start within SEQ ID NO 4208: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4210
- Ceres seq_id 1579913
- Location of start within SEQ ID NO 4208: at 157 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1828
- gi No. 3169719
- Description: (AF007109) similar to yeast dcp1 [Arabidopsis thaliana]
- % Identity: 71.4
- Alignment Length: 84
- Location of Alignment in SEQ ID NO 4210: from 11 to 93

Maximum Length Sequence:
related to:
Clone IDs:
341409
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4211
- Ceres seq_id 1579922
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4212
- Ceres seq_id 1579923
- Location of start within SEQ ID NO 4211: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1829
- gi No. 2826786
- Description: (Y10905) RAPB protein [Oryza sativa]
- % Identity: 76.5
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 4212: from 47 to 164

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4213
- Ceres seq_id 1579924
- Location of start within SEQ ID NO 4211: at 139 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1830
- gi No. 2826786
- Description: (Y10905) RAPB protein [Oryza sativa]
- % Identity: 76.5
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 4213: from 1 to 118

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4214
- Ceres seq_id 1579925
- Location of start within SEQ ID NO 4211: at 175 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1831
- gi No. 2826786
- Description: (Y10905) RAPB protein [Oryza sativa]
- % Identity: 76.5
- Alignment Length: 119
- Location of Alignment in SEQ ID NO 4214: from 1 to 106

Maximum Length Sequence:
related to:
Clone IDs:
341436

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 4215
- Ceres seq_id 1579926

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4216
- Ceres seq_id 1579927
- Location of start within SEQ ID NO 4215: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

- Helix-loop-helix DNA-binding domain
- Location within SEQ ID NO 4216: from 53 to 98 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341452

(Ac) cDNA Polynucleotide Sequence

- Pat. Appln. SEQ ID NO 4217
- Ceres seq_id 1579931

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4218
- Ceres seq_id 1579932
- Location of start within SEQ ID NO 4217: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence

- Pat. Appln. SEQ ID NO 4219
- Ceres seq_id 1579933
- Location of start within SEQ ID NO 4217: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

- Alignment No. 1832
- gi No. 1079136
- Description: RNA-binding protein cabeza - fruit fly (Drosophila melanogaster) >gi|532788 (U13178) RNA binding protein [Drosophila

melanogaster] >gi|567106 (L37083) RNA binding protein [Drosophila melanogaster]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4219: from 42 to 52

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4220
- Ceres seq_id 1579934
- Location of start within SEQ ID NO 4217: at 123 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1833
- gi No. 1079136
- Description: RNA-binding protein cabeza - fruit fly (Drosophila melanogaster) >gi|532788 (U13178) RNA binding protein [Drosophila melanogaster] >gi|567106 (L37083) RNA binding protein [Drosophila melanogaster]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4220: from 2 to 12

Maximum Length Sequence:
related to:
Clone IDs:
341528

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4221
- Ceres seq_id 1579953

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4222
- Ceres seq_id 1579954
- Location of start within SEQ ID NO 4221: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4222: from 82 to 131 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1834
- gi No. 1079080
- Description: GCR 20 protein - fruit fly (Drosophila melanogaster) >gi|510504|emb|CAA51651| (X73134) GCR 20 [Drosophila melanogaster]
- % Identity: 70
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 4222: from 83 to 112

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4223
- Ceres seq_id 1579955
- Location of start within SEQ ID NO 4221: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1835
- gi No. 2754696

- Description: (D88440) high molecular mass nuclear antigen [Gallus gallus]
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4223: from 27 to 38

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4224
- Ceres seq_id 1579956
- Location of start within SEQ ID NO 4221: at 34 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4224: from 71 to 120 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1836
- gi No. 1079080
- Description: GCR 20 protein - fruit fly (Drosophila melanogaster) >gi|510504|emb|CAA51651| (X73134) GCR 20 [Drosophila melanogaster]
- % Identity: 70
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 4224: from 72 to 101

Maximum Length Sequence:
related to:
Clone IDs:
341615
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4225
- Ceres seq_id 1579968

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4226
- Ceres seq_id 1579969
- Location of start within SEQ ID NO 4225: at 177 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1837
- gi No. 2281633
- Description: (AF003097) AP2 domain containing protein RAP2.4 [Arabidopsis thaliana]
- % Identity: 72.2
- Alignment Length: 36
- Location of Alignment in SEQ ID NO 4226: from 72 to 105

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4227
- Ceres seq_id 1579970
- Location of start within SEQ ID NO 4225: at 222 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1838
- gi No. 2281633
- Description: (AF003097) AP2 domain containing protein RAP2.4 [Arabidopsis thaliana]
- % Identity: 72.2

- Alignment Length: 36
- Location of Alignment in SEQ ID NO 4227: from 57 to 90

Maximum Length Sequence:
related to:
Clone IDs:
341639
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4228
- Ceres seq_id 1579975
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4229
- Ceres seq_id 1579976
- Location of start within SEQ ID NO 4228: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1839
- gi No. 2134209
- Description: protamine II-3 - painted turtle
- % Identity: 75
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4229: from 56 to 67

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4230
- Ceres seq_id 1579977
- Location of start within SEQ ID NO 4228: at 117 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1840
- gi No. 2134209
- Description: protamine II-3 - painted turtle
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4230: from 49 to 59

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4231
- Ceres seq_id 1579978
- Location of start within SEQ ID NO 4228: at 129 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1841
- gi No. 2134209
- Description: protamine II-3 - painted turtle
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4231: from 45 to 55

Maximum Length Sequence:
related to:
Clone IDs:
341657
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4232

- Ceres seq_id 1579979

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4233
- Ceres seq_id 1579980
- Location of start within SEQ ID NO 4232: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4233: from 2 to 89 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4234
- Ceres seq_id 1579981
- Location of start within SEQ ID NO 4232: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341668
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4235
- Ceres seq_id 1579986

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4236
- Ceres seq_id 1579987
- Location of start within SEQ ID NO 4235: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Formate--tetrahydrofolate ligase
- Location within SEQ ID NO 4236: from 1 to 159 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1842
- gi No. 2507455
- Description: FORMATE--TETRAHYDROFOLATE LIGASE (FORMYLTETRAHYDROFOLATE SYNTHETASE) (FHS) (FTHFS) >gi|322401|pir||A43350 formate--tetrahydrofolate ligase (EC 6.3.4.3) - spinach >gi|170145 (M83940) 10-formyltetrahydrofolate synthetase [Spinacia oleracea]
- % Identity: 78.8
- Alignment Length: 160
- Location of Alignment in SEQ ID NO 4236: from 1 to 159

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4237
- Ceres seq_id 1579988
- Location of start within SEQ ID NO 4235: at 66 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Formate--tetrahydrofolate ligase
- Location within SEQ ID NO 4237: from 1 to 138 aa.

(Dp) Related Amino Acid Sequences

- Alignment No. 1843
- gi No. 2507455
- Description: FORMATE--TETRAHYDROFOLATE LIGASE (FORMYLTETRAHYDROFOLATE SYNTHETASE) (FHS) (FTHFS) >gi|322401|pir||A43350 formate--tetrahydrofolate ligase (EC 6.3.4.3) - spinach >gi|170145 (M83940) 10-formyltetrahydrofolate synthetase [Spinacia oleracea]
- % Identity: 78.8
- Alignment Length: 160
- Location of Alignment in SEQ ID NO 4237: from 1 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4238
- Ceres seq_id 1579989
- Location of start within SEQ ID NO 4235: at 90 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Formate--tetrahydrofolate ligase
- Location within SEQ ID NO 4238: from 1 to 130 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1844
- gi No. 2507455
- Description: FORMATE--TETRAHYDROFOLATE LIGASE (FORMYLTETRAHYDROFOLATE SYNTHETASE) (FHS) (FTHFS) >gi|322401|pir||A43350 formate--tetrahydrofolate ligase (EC 6.3.4.3) - spinach >gi|170145 (M83940) 10-formyltetrahydrofolate synthetase [Spinacia oleracea]
- % Identity: 78.8
- Alignment Length: 160
- Location of Alignment in SEQ ID NO 4238: from 1 to 130

Maximum Length Sequence:
related to:
Clone IDs:
341676
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4239
- Ceres seq_id 1579994

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4240
- Ceres seq_id 1579995
- Location of start within SEQ ID NO 4239: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 4240: from 36 to 143 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1845
- gi No. 443954
- Description: (Z14081) cyclophilin [Nicotiana tabacum]
- % Identity: 73.9
- Alignment Length: 23
- Location of Alignment in SEQ ID NO 4240: from 118 to 140

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4241
- Ceres seq_id 1579996
- Location of start within SEQ ID NO 4239: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Peptidyl-prolyl cis-trans isomerase
- Location within SEQ ID NO 4241: from 3 to 110 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1846
- gi No. 443954
- Description: (Z14081) cyclophilin [Nicotiana tabacum]
- % Identity: 73.9
- Alignment Length: 23
- Location of Alignment in SEQ ID NO 4241: from 85 to 107

Maximum Length Sequence:
related to:
Clone IDs:
341866
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4242
- Ceres seq_id 1580028
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4243
- Ceres seq_id 1580029
- Location of start within SEQ ID NO 4242: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4243: from 10 to 94 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341869
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4244
- Ceres seq_id 1580030
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4245
- Ceres seq_id 1580031
- Location of start within SEQ ID NO 4244: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1847
- gi No. 4836401
- Description: (AF118023) SH3 domain-binding protein SNP70 [Homo sapiens]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4245: from 14 to 24

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4246
- Ceres seq_id 1580032
- Location of start within SEQ ID NO 4244: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1848
- gi No. 102426
- Description: collagen-related protein 3 precursor - Hydra magnipapillata >gi|102430|pir||S21931 mini-collagen - Hydra sp >gi|9451|emb|CAA43381| (X61047) mini-collagen [Hydra sp.]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4246: from 3 to 13

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4247
- Ceres seq_id 1580033
- Location of start within SEQ ID NO 4244: at 17 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341952
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4248
- Ceres seq_id 1580051
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4249
- Ceres seq_id 1580052
- Location of start within SEQ ID NO 4248: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4250
- Ceres seq_id 1580053
- Location of start within SEQ ID NO 4248: at 96 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences
- Alignment No. 1849
- gi No. 464705
- Description: 40S RIBOSOMAL PROTEIN S13 >gi|419802|pir||S30146 ribosomal protein S13.e - maize >gi|288059|emb|CAA44311| (X62455) cytoplasmatic ribosomal protein S13 [Zea mays]
- % Identity: 77.4
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 4250: from 1 to 31

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4251
- Ceres seq_id 1580054
- Location of start within SEQ ID NO 4248: at 105 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1850
- gi No. 464705
- Description: 40S RIBOSOMAL PROTEIN S13 >gi|419802|pir||S30146 ribosomal protein S13.e - maize >gi|288059|emb|CAA44311| (X62455) cytoplasmatic ribosomal protein S13 [Zea mays]
- % Identity: 77.4
- Alignment Length: 31
- Location of Alignment in SEQ ID NO 4251: from 1 to 28

Maximum Length Sequence:
related to:
Clone IDs:
341958
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4252
- Ceres seq_id 1580058
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4253
- Ceres seq_id 1580059
- Location of start within SEQ ID NO 4252: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- 7 transmembrane receptor (rhodopsin family)
- Location within SEQ ID NO 4253: from 10 to 71 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4254
- Ceres seq_id 1580060
- Location of start within SEQ ID NO 4252: at 106 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341959
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4255
- Ceres seq_id 1580061
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4256
- Ceres seq_id 1580062
- Location of start within SEQ ID NO 4255: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 4256: from 116 to 146 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4257
- Ceres seq_id 1580063
- Location of start within SEQ ID NO 4255: at 126 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Cytidine and deoxycytidylate deaminase zinc-binding region
- Location within SEQ ID NO 4257: from 75 to 105 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341990
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4258
- Ceres seq_id 1580068
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4259
- Ceres seq_id 1580069
- Location of start within SEQ ID NO 4258: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Galactose-1-phosphate uridyl transferase
- Location within SEQ ID NO 4259: from 20 to 79 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4260
- Ceres seq_id 1580070
- Location of start within SEQ ID NO 4258: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4261
- Ceres seq_id 1580071
- Location of start within SEQ ID NO 4258: at 104 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
341995
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4262
- Ceres seq_id 1580072
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4263
- Ceres seq_id 1580073
- Location of start within SEQ ID NO 4262: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 4263: from 33 to 131 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4264
- Ceres seq_id 1580074
- Location of start within SEQ ID NO 4262: at 166 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Glycosyl hydrolases family 17
- Location within SEQ ID NO 4264: from 1 to 76 aa.

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
342041
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4265
- Ceres seq_id 1580088
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4266
- Ceres seq_id 1580089
- Location of start within SEQ ID NO 4265: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic porin
- Location within SEQ ID NO 4266: from 36 to 138 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1851
- gi No. 1172553
- Description: OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN (VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN) (VDAC) >gi|456672|emb|CAA54788|(X77733) voltage dependent anion channel (VDAC) [Triticum aestivum]
- % Identity: 80
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 4266: from 34 to 138

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4267
- Ceres seq_id 1580090
- Location of start within SEQ ID NO 4265: at 101 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic porin
- Location within SEQ ID NO 4267: from 3 to 105 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1852
- gi No. 1172553
- Description: OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN (VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN) (VDAC) >gi|456672|emb|CAA54788|(X77733) voltage dependent anion channel (VDAC) [Triticum aestivum]
- % Identity: 80
- Alignment Length: 105
- Location of Alignment in SEQ ID NO 4267: from 1 to 105

Maximum Length Sequence:
related to:
Clone IDs:
342117
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4268
- Ceres seq_id 1580112
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4269
- Ceres seq_id 1580113
- Location of start within SEQ ID NO 4268: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4270
- Ceres seq_id 1580114
- Location of start within SEQ ID NO 4268: at 160 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Histone-like transcription factors (CBF/NF-Y) and archaeal histones.
- Location within SEQ ID NO 4270: from 32 to 100 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1853
- gi No. 2398529
- Description: (Y13724) Transcription factor [Arabidopsis thaliana]
- % Identity: 78
- Alignment Length: 82
- Location of Alignment in SEQ ID NO 4270: from 28 to 108

Maximum Length Sequence:
related to:
Clone IDs:
342118
(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4271
- Ceres seq_id 1580115
(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4272
- Ceres seq_id 1580116
- Location of start within SEQ ID NO 4271: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Collagen triple helix repeat (20 copies)
- Location within SEQ ID NO 4272: from 28 to 74 aa.

(Dp) Related Amino Acid Sequences
- Alignment No. 1854
- gi No. 1185397
- Description: (U25281) SH3 domain binding protein [Rattus norvegicus] >gi|1587070|prf||2205340A CR16 gene [Rattus norvegicus]
- % Identity: 83.3
- Alignment Length: 12
- Location of Alignment in SEQ ID NO 4272: from 64 to 75

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4273
- Ceres seq_id 1580117
- Location of start within SEQ ID NO 4271: at 2 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1855
- gi No. 585296
- Description: INFECTED CELL PROTEIN ICP34.5 (NEUROVIRULENCE FACTOR ICP34.5) >gi|330119 (M33701) infected-cell protein 34.5 [Herpes simplex virus type 1]
- % Identity: 72.7
- Alignment Length: 11
- Location of Alignment in SEQ ID NO 4273: from 2 to 12

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4274
- Ceres seq_id 1580118
- Location of start within SEQ ID NO 4271: at 59 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences

Maximum Length Sequence:
related to:
Clone IDs:
342123

(Ac) cDNA Polynucleotide Sequence
- Pat. Appln. SEQ ID NO 4275
- Ceres seq_id 1580119

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4276
- Ceres seq_id 1580120
- Location of start within SEQ ID NO 4275: at 1 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)
- Eukaryotic protein kinase domain
- Location within SEQ ID NO 4276: from 2 to 95 aa.

(Dp) Related Amino Acid Sequences

(B) Polypeptide Sequence
- Pat. Appln. SEQ ID NO 4277
- Ceres seq_id 1580121
- Location of start within SEQ ID NO 4275: at 3 nt.

(C) Nomination and Annotation of Domains within Predicted Polypeptide(s)

(Dp) Related Amino Acid Sequences
- Alignment No. 1856
- gi No. 2252864
- Description: (AF013294) Similar to receptor kinase [Arabidopsis thaliana]
- % Identity: 70
- Alignment Length: 30
- Location of Alignment in SEQ ID NO 4277: from 94 to 123